



凸性与凸优化（合订版）

Combined Edition · Chapters 2–4

梁经纬 Jingwei Liang

Institute of Natural Sciences
School of Mathematical Sciences

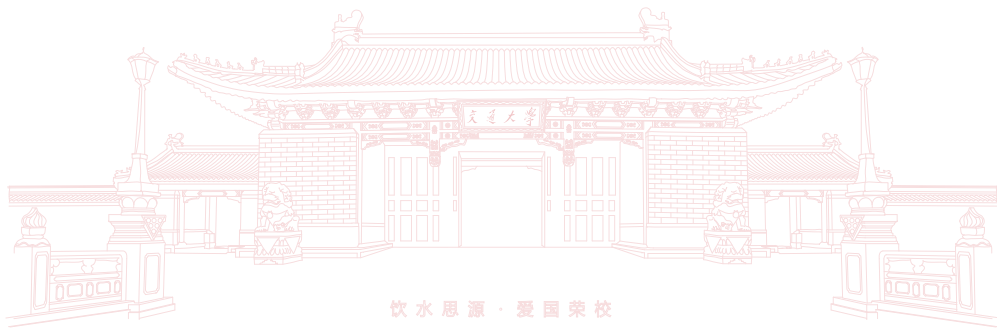
Lecture 01: 凸性与凸优化（合订版）

2026 年秋季学期

饮水思源 · 爱国荣校

01

优化与凸性



生产计划

一家工厂在一个班次内生产 A、B 两类产品，产量为

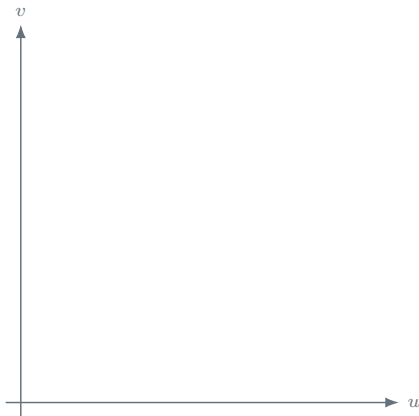
$$x = \begin{pmatrix} u \\ v \end{pmatrix} \in \mathbb{R}^2.$$

两道工序的单位工时与班次容量为

	A	B	容量
工序 1	2	1	8
工序 2	1	2	8

每批收益分别为 3 与 2，因此模型为

$$\begin{aligned} & \underset{u,v}{\text{minimize}} && -3u - 2v \\ & \text{subject to} && 2u + v \leq 8, \\ & && u + 2v \leq 8, \\ & && u \geq 0, \quad v \geq 0. \end{aligned}$$



生产计划

一家工厂在一个班次内生产 A、B 两类产品，产量为

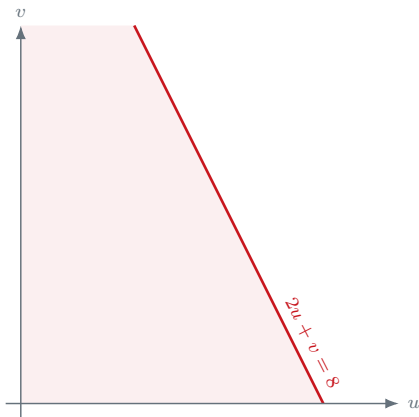
$$x = \begin{pmatrix} u \\ v \end{pmatrix} \in \mathbb{R}^2.$$

两道工序的单位工时与班次容量为

	A	B	容量
工序 1	2	1	8
工序 2	1	2	8

每批收益分别为 3 与 2，因此模型为

$$\begin{aligned} & \underset{u,v}{\text{minimize}} && -3u - 2v \\ & \text{subject to} && 2u + v \leq 8, \\ & && u + 2v \leq 8, \\ & && u \geq 0, \quad v \geq 0. \end{aligned}$$



生产计划

一家工厂在一个班次内生产 A、B 两类产品，产量为

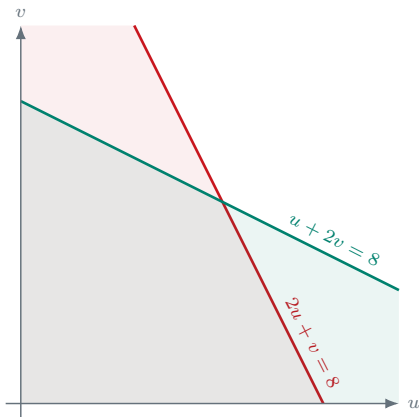
$$x = \begin{pmatrix} u \\ v \end{pmatrix} \in \mathbb{R}^2.$$

两道工序的单位工时与班次容量为

	A	B	容量
工序 1	2	1	8
工序 2	1	2	8

每批收益分别为 3 与 2，因此模型为

$$\begin{aligned} & \underset{u,v}{\text{minimize}} && -3u - 2v \\ & \text{subject to} && 2u + v \leq 8, \\ & && u + 2v \leq 8, \\ & && u \geq 0, \quad v \geq 0. \end{aligned}$$



生产计划

一家工厂在一个班次内生产 A、B 两类产品，产量为

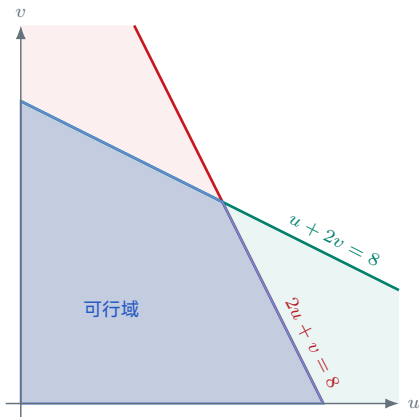
$$x = \begin{pmatrix} u \\ v \end{pmatrix} \in \mathbb{R}^2.$$

两道工序的单位工时与班次容量为

	A	B	容量
工序 1	2	1	8
工序 2	1	2	8

每批收益分别为 3 与 2，因此模型为

$$\begin{aligned} & \underset{u,v}{\text{minimize}} && -3u - 2v \\ & \text{subject to} && 2u + v \leq 8, \\ & && u + 2v \leq 8, \\ & && u \geq 0, \quad v \geq 0. \end{aligned}$$



生产计划

一家工厂在一个班次内生产 A、B 两类产品，产量为

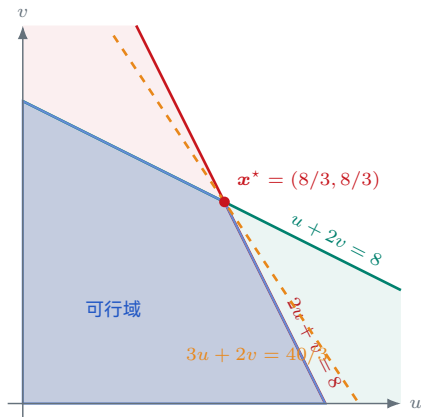
$$x = \begin{pmatrix} u \\ v \end{pmatrix} \in \mathbb{R}^2.$$

两道工序的单位工时与班次容量为

	A	B	容量
工序 1	2	1	8
工序 2	1	2	8

每批收益分别为 3 与 2，因此模型为

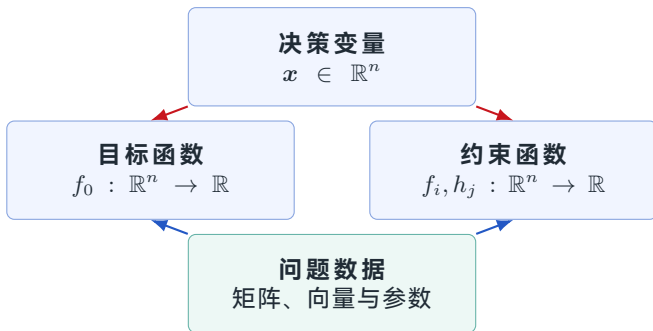
$$\begin{aligned} & \underset{u,v}{\text{minimize}} && -3u - 2v \\ & \text{subject to} && 2u + v \leq 8, \\ & && u + 2v \leq 8, \\ & && u \geq 0, \quad v \geq 0. \end{aligned}$$



自编示例；参见 Boyd & Vandenberghe, Section 4.3.1

优化问题的组成

$$\begin{aligned} & \underset{x \in \mathbb{R}^n}{\text{minimize}} && f_0(x) \\ & \text{subject to} && f_i(x) \leq 0, \quad i = 1, \dots, m, \\ & && h_j(x) = 0, \quad j = 1, \dots, p. \end{aligned}$$

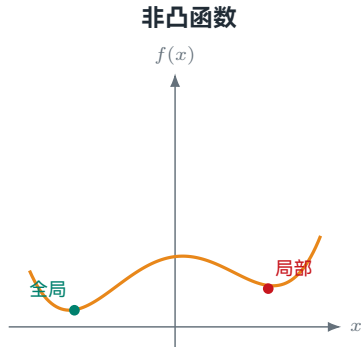
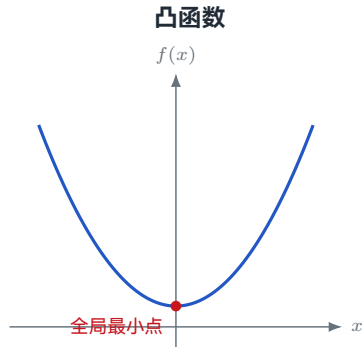


在最小二乘模型中,

$$A \in \mathbb{R}^{m \times n}, \quad x \in \mathbb{R}^n, \quad b \in \mathbb{R}^m, \quad Ax - b \in \mathbb{R}^m.$$

维度决定乘法是否成立, 也决定梯度与约束所在的空间。

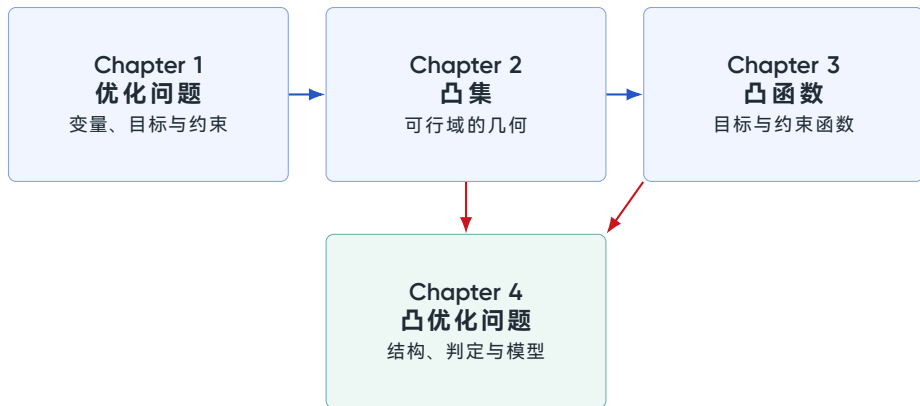
凸性带来的结构



凸性的好处:

- 若目标函数与可行域满足凸性条件, 则局部最优点也是全局最优点;
- 一阶信息可以给出全局最优性判定。凸性并不保证问题规模小, 但提供了可验证的结构。

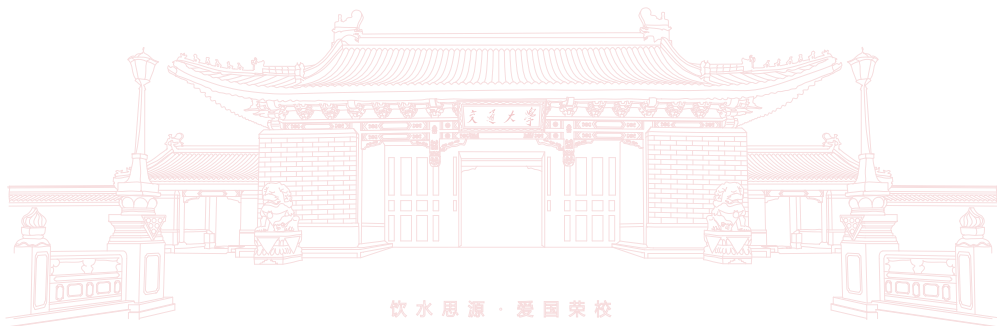
前四章的关系



本讲保留这一知识顺序。集合回答“允许哪些点”，函数回答“如何比较这些点”，二者共同决定优化问题是否为凸优化问题。

02

凸集

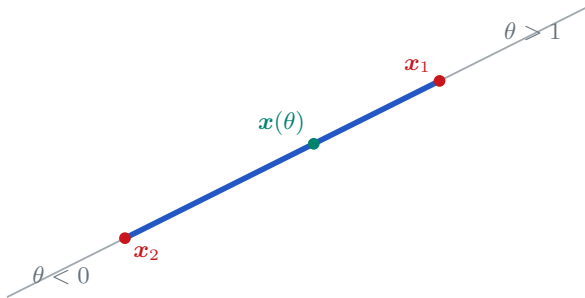


饮水思源 · 爱国荣校

直线与线段

给定 $x_1 \neq x_2 \in \mathbb{R}^n$, 经过两点的直线可参数化为

$$x(\theta) = \theta x_1 + (1 - \theta)x_2 = x_2 + \theta(x_1 - x_2), \quad \theta \in \mathbb{R}.$$



$\theta = 0 \Rightarrow x_2$, $\theta = 1 \Rightarrow x_1$, $0 \leq \theta \leq 1 \Rightarrow x(\theta)$ 位于两点间的线段。

仿射集与仿射组合

集合 $\mathcal{A} \subseteq \mathbb{R}^n$ 称为**仿射集** (affine set), 若

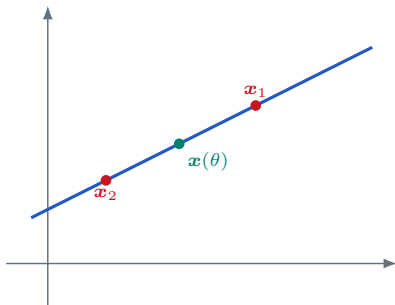
$$x_1, x_2 \in \mathcal{A}, \quad \theta \in \mathbb{R} \implies \theta x_1 + (1 - \theta)x_2 \in \mathcal{A}.$$

也就是说, 任意两点所在的整条直线都留在 $\mathcal{A}$ 内。

更一般地, 若

$$x = \sum_{i=1}^k \theta_i x_i, \quad \sum_{i=1}^k \theta_i = 1,$$

则称 x 为 $x_1, \dots, x_k$ 的**仿射组合** (affine combination)。系数可以为负。



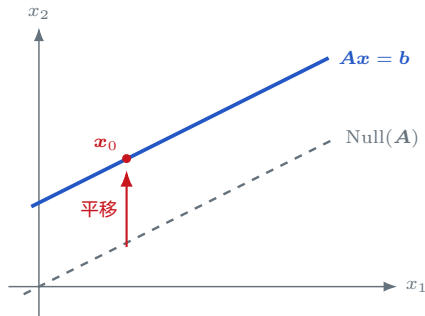
线性方程组的解集

若 $\mathcal{A} = \{x \mid Ax = b\}$ 非空, 则 $\mathcal{A}$ 是仿射集:

$$A(\theta x_1 + (1 - \theta)x_2) = \theta b + (1 - \theta)b = b.$$

取任意一个可行点 x_0 , 则

$$\boxed{\mathcal{A} = x_0 + \text{Null}(A)} = \{x_0 + v \mid Av = 0\}.$$



仿射集是“线性子空间加一个偏移量”; 偏移量由 b 决定。

Boyd & Vandenberghe, Section 2.1.2, Example 2.1

仿射包与仿射维数

集合 $\mathcal{C}$ 的**仿射包** (affine hull) 是包含它的最小仿射集:

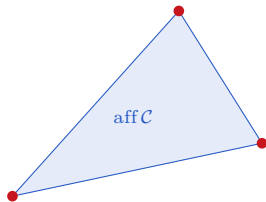
$$\text{aff } \mathcal{C} = \left\{ \sum_{i=1}^k \theta_i \mathbf{x}_i \mid \mathbf{x}_i \in \mathcal{C}, \sum_{i=1}^k \theta_i = 1 \right\}.$$

若 $\mathcal{S}$ 是仿射集且 $\mathcal{C} \subseteq \mathcal{S}$, 则 $\text{aff } \mathcal{C} \subseteq \mathcal{S}$ 。

仿射维数定义为

$$\dim_{\text{aff}} \mathcal{C} = \dim(\text{aff } \mathcal{C}).$$

三个不共线点的仿射包是一个平面; 一条非退化线段的仿射维数为 1。



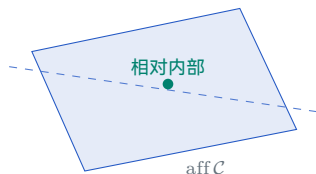
相对内部

当 C 落在低维仿射集内时，普通内部可能为空。相对内部在 $\text{aff } C$ 中取邻域：

$$\text{relint } C = \{x \in C \mid \exists r > 0, B(x, r) \cap \text{aff } C \subseteq C\}.$$

例如在 $\mathbb{R}^3$ 中，平面 $x_3 = 0$ 上的实心方形普通内部为空，但其相对内部是去掉边界后的方形。相对边界为

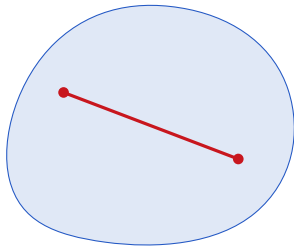
$$\text{relbd } C = \text{cl } C \setminus \text{relint } C.$$



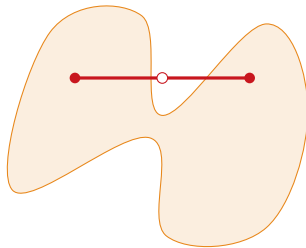
凸集

集合 $C \subseteq \mathbb{R}^n$ 是**凸集** (convex set), 若

$$x, y \in C, \quad 0 \leq \theta \leq 1 \quad \implies \quad \theta x + (1 - \theta)y \in C.$$



凸集



非凸集

仿射集包含整条直线，因而必为凸集；凸集只要求连接两点的线段留在集合内。

凸组合与凸包

若

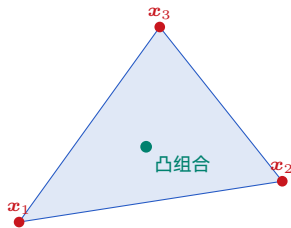
$$\mathbf{x} = \sum_{i=1}^k \theta_i \mathbf{x}_i, \quad \theta_i \geq 0, \quad \sum_{i=1}^k \theta_i = 1,$$

则 $\mathbf{x}$ 是这些点的**凸组合** (convex combination)。它可理解为非负权重的混合。

所有凸组合构成**凸包** (convex hull):

$$\text{conv } \mathcal{C} = \left\{ \sum_{i=1}^k \theta_i \mathbf{x}_i \mid \mathbf{x}_i \in \mathcal{C}, \theta_i \geq 0, \sum_i \theta_i = 1 \right\}.$$

$\text{conv } \mathcal{C}$ 是包含 $\mathcal{C}$ 的最小凸集。



平均与期望

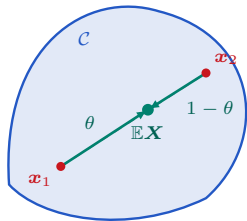
凸性不仅控制有限个点的加权平均，也控制极限、积分和期望。

若 C 为凸集, $\theta_i \geq 0$, $\sum_{i=1}^{\infty} \theta_i = 1$, $x_i \in C$, 且级数收敛, 则

$$\sum_{i=1}^{\infty} \theta_i x_i \in C.$$

若随机向量 X 以概率 1 落在 C , 且期望存在, 则

$$\mathbb{E}X \in C.$$



概率质量的重心仍在 C 内

二点分布已经给出通常的凸组合:

$$\begin{aligned} \mathbb{P}(X = x_1) &= \theta, \quad \mathbb{P}(X = x_2) = 1 - \theta \\ \implies \mathbb{E}X &= \theta x_1 + (1 - \theta)x_2. \end{aligned}$$

锥与锥包

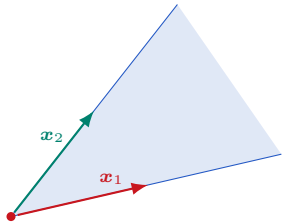
集合 $\mathcal{K}$ 为**锥** (cone), 若

$$x \in \mathcal{K}, \quad \alpha \geq 0 \quad \implies \quad \alpha x \in \mathcal{K}.$$

若它还满足凸性, 则

$$x_1, x_2 \in \mathcal{K}, \quad \alpha_1, \alpha_2 \geq 0 \quad \implies \quad \alpha_1 x_1 + \alpha_2 x_2 \in \mathcal{K}.$$

$\sum_i \alpha_i x_i, \alpha_i \geq 0$ 称为**锥组合**; 所有锥组合构成 $\text{cone } \mathcal{C}$, 即包含 $\mathcal{C}$ 的最小凸锥。



基本例子

集合	凸性/仿射性	锥性
$\emptyset, \{x_0\}, \mathbb{R}^n$	仿射，因此凸	仅 $\emptyset, \{0\}, \mathbb{R}^n$ 为锥
直线	仿射	经过原点时是子空间、凸锥
线段	凸，通常非仿射	仅退化为原点时为锥
射线 $\{x_0 + \theta v \mid \theta \geq 0\}$	凸，通常非仿射	$x_0 = 0$ 时为凸锥
线性子空间	仿射，因此凸	凸锥

这些基本对象是后续构造可行域的积木。区分“仿射”“凸”“锥”时，应分别检查：任意实系数的仿射组合、非负且和为一的凸组合、非负缩放。

超平面

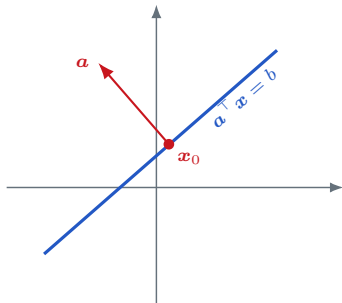
给定 $\mathbf{a} \in \mathbb{R}^n \setminus \{\mathbf{0}\}$ 、 $b \in \mathbb{R}$,

$$\mathcal{H} = \{\mathbf{x} \mid \mathbf{a}^\top \mathbf{x} = b\}$$

称为**超平面** (hyperplane)。若 $\mathbf{x}_0 \in \mathcal{H}$, 则

$$\mathcal{H} = \{\mathbf{x} \mid \mathbf{a}^\top (\mathbf{x} - \mathbf{x}_0) = 0\} = \mathbf{x}_0 + \mathbf{a}^\perp.$$

$\mathbf{a}$ 是法向量：在超平面内移动的任何向量 $\mathbf{v}$ 都满足 $\mathbf{a}^\top \mathbf{v} = 0$ 。常数 b 决定超平面距原点的偏移。



半空间

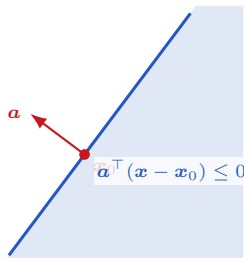
与超平面对应的闭半空间为

$$\mathcal{H}^- = \{x \mid a^\top x \leq b\} = \{x \mid a^\top (x - x_0) \leq 0\}.$$

边界是 $\{x \mid a^\top x = b\}$ ，内部是开半空间 $\{x \mid a^\top x < b\}$ 。

若 $a^\top (x - x_0) \leq 0$ ，则 $x - x_0$ 与外法向量 a 的夹角为直角或钝角。

半空间是凸集，因为线性不等式在凸组合下保持。



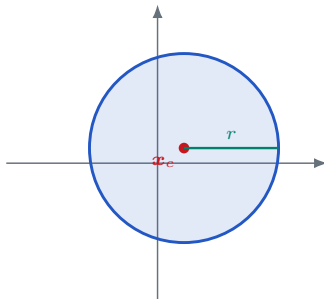
欧氏球

中心 x_c 、半径 $r > 0$ 的欧氏球为

$$B(x_c, r) = \{x \mid \|x - x_c\|_2 \leq r\} = \{x \mid (x - x_c)^\top (x - x_c) \leq r^2\}.$$

它是凸集：若 $x_1, x_2 \in B(x_c, r)$ ，则对 $0 \leq \theta \leq 1$,

$$\begin{aligned} \|\theta x_1 + (1 - \theta)x_2 - x_c\|_2 &\leq \theta \|x_1 - x_c\|_2 + (1 - \theta) \|x_2 - x_c\|_2 \\ &\leq r. \end{aligned}$$



椭球

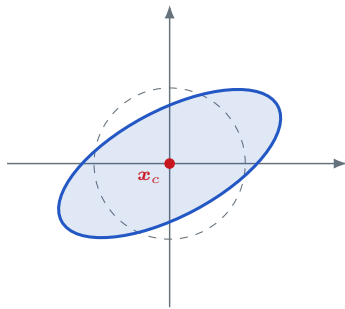
若 $P = P^\top \succ 0$, 椭球定义为

$$\mathcal{E} = \{x \mid (x - x_c)^\top P^{-1}(x - x_c) \leq 1\}.$$

令 $P = QQ^\top$ (此时 Q 可逆), 则同一集合也可写作

$$\mathcal{E} = \{x_c + Qu \mid \|u\|_2 \leq 1\}.$$

它是单位欧氏球的仿射像。若 λ_i 是 P 的特征值, 则半轴长度为 $\sqrt{\lambda_i}$ 。更一般地, $\{x_c + Qu \mid \|u\|_2 \leq 1\}$ 允许 Q 奇异; 此时称为退化椭球。



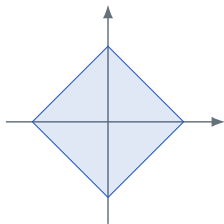
Boyd & Vandenberghe, Section 2.2.2

范数球

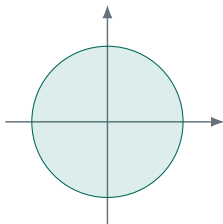
任意范数 $\|\cdot\|$ 的球

$$\{\mathbf{x} \mid \|\mathbf{x} - \mathbf{x}_c\| \leq r\}$$

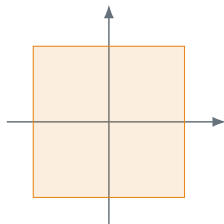
都是凸集。二维单位球说明范数选择改变几何形状，但不改变凸性。



$$\|\mathbf{x}\|_1 \leq 1$$



$$\|\mathbf{x}\|_2 \leq 1$$



$$\|\mathbf{x}\|_\infty \leq 1$$

$$\|\mathbf{x}\|_1 = \sum_i |x_i|, \quad \|\mathbf{x}\|_2 = (\mathbf{x}^\top \mathbf{x})^{1/2}, \quad \|\mathbf{x}\|_\infty = \max_i |x_i|.$$

范数锥与二阶锥

任意范数诱导一个**范数锥** (norm cone)

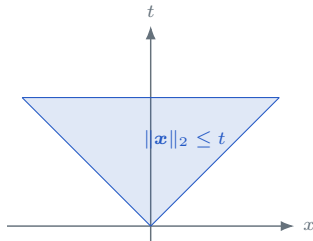
$$\mathcal{K}_{\|\cdot\|} = \{(\mathbf{x}, t) \in \mathbb{R}^{n+1} \mid \|\mathbf{x}\| \leq t\},$$

它是凸锥。欧氏范数给出二阶锥 (second-order cone, SOC):

$$\mathcal{Q}^{n+1} = \{(\mathbf{x}, t) \mid \|\mathbf{x}\|_2 \leq t\}.$$

$$\|\mathbf{x}\|_2 \leq t \iff \mathbf{x}^\top \mathbf{x} \leq t^2, \quad t \geq 0.$$

它也称 Lorentz cone 或 quadratic cone。



Boyd & Vandenberghe, Section 2.2.3, Example 2.3

多面体

有限个线性等式和不等式的公共解集称为**多面体** (polyhedron):

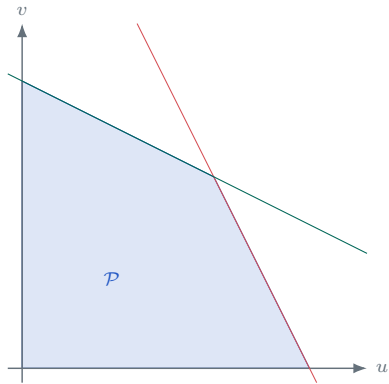
$$\mathcal{P} = \{x \in \mathbb{R}^n \mid Ax \preceq b, Cx = d\}.$$

它是半空间和超平面的有限交集，因而是凸集。

二维生产约束

$$2u + v \leq 8, \quad u + 2v \leq 8, \quad u, v \geq 0$$

定义一个有界多面体；有界多面体常称为 polytope。



Boyd & Vandenberghe, Section 2.2.4

非负正交锥

非负正交锥 (nonnegative orthant) 为

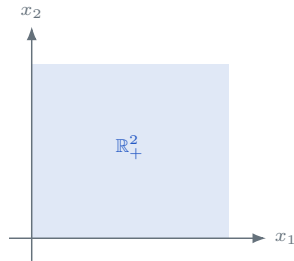
$$\mathbb{R}_+^n = \{\mathbf{x} \in \mathbb{R}^n \mid x_i \geq 0, i = 1, \dots, n\} = \{\mathbf{x} \in \mathbb{R}^n \mid \mathbf{x} \succeq \mathbf{0}\}.$$

它同时是多面体和凸锥。分量不等式 $\mathbf{x} \preceq \mathbf{y}$ 表示每个分量均满足 $x_i \leq y_i$ 。

例如

$$\{\mathbf{x} \mid \mathbf{Ax} \preceq \mathbf{b}, \mathbf{Cx} = \mathbf{d}\} = \{\mathbf{x} \mid (\mathbf{b} - \mathbf{Ax}, \mathbf{d} - \mathbf{Cx}) \in \mathbb{R}_+^m \times \{\mathbf{0}\}\}.$$

这把线性约束表达为一个仿射映射的原像。



Boyd & Vandenberghe, Sections 2.2.4, 2.3.2, and 2.4, Examples 2.4, 2.9, and 2.14

单纯形

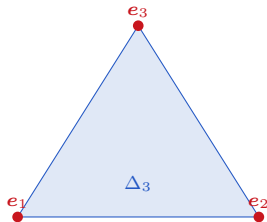
由仿射无关的 $k + 1$ 个点 $v_0, \dots, v_k$ 决定的 k 维单纯形为

$$\text{conv}\{v_0, \dots, v_k\} = \left\{ \sum_{i=0}^k \theta_i v_i \mid \theta \succeq \mathbf{0}, \mathbf{1}^\top \theta = 1 \right\}.$$

一维单纯形是线段，二维单纯形是实心三角形，三维单纯形是实心四面体。

单位单纯形: $x \succeq 0, \mathbf{1}^\top x \leq 1$ 。

概率单纯形: $x \succeq 0, \mathbf{1}^\top x = 1$ 。



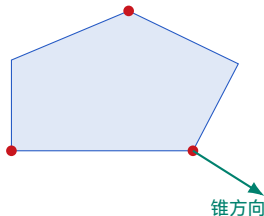
Boyd & Vandenberghe, Section 2.2.4, Example 2.5

多面体的两种描述

多面体可由线性约束描述（H-表示），也可由有限个生成点和方向描述：

$$\left\{ \sum_{i=1}^m \theta_i v_i + \sum_{i=m+1}^k \alpha_i v_i \mid \theta \succeq 0, \mathbf{1}^\top \theta = 1, \alpha \succeq 0 \right\}.$$

前一部分是 $v_1, \dots, v_m$ 的凸包，后一部分是余下向量的锥包。每个多面体都存在这种表示；它与 H-表示在建模和计算中的复杂度可能相差很大。



半正定锥

记 $\mathbb{S}^n = \{\mathbf{X} \in \mathbb{R}^{n \times n} \mid \mathbf{X} = \mathbf{X}^\top\}$ 。半正定锥为

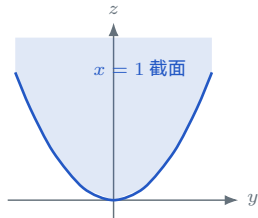
$$\mathbb{S}_+^n = \{\mathbf{X} \in \mathbb{S}^n \mid \mathbf{z}^\top \mathbf{X} \mathbf{z} \geq 0, \forall \mathbf{z} \in \mathbb{R}^n\}.$$

严格不等式 $\mathbf{X} \succ 0$ 对应正定锥 $\mathbb{S}_{++}^n$ 。

$$\mathbf{X}, \mathbf{Y} \succeq 0, \quad \alpha, \beta \geq 0 \quad \Rightarrow \quad \mathbf{z}^\top (\alpha \mathbf{X} + \beta \mathbf{Y}) \mathbf{z} \geq 0.$$

故 $\mathbb{S}_+^n$ 是凸锥。对于 $\mathbf{X} = \begin{bmatrix} x & y \\ y & z \end{bmatrix}$,

$$\mathbf{X} \succeq 0 \iff x \geq 0, z \geq 0, xz \geq y^2.$$



交集保持凸性

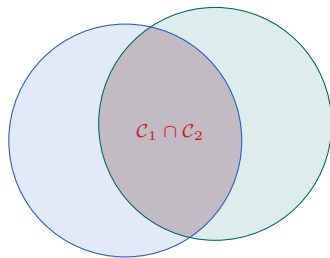
若每个 $\mathcal{C}_\alpha$ 都是凸集，则任意交集仍是凸集：

$$\bigcap_{\alpha \in \mathcal{I}} \mathcal{C}_\alpha \text{ 为凸集.}$$

事实上，若 x, y 同时属于所有 $\mathcal{C}_\alpha$ ，则它们的凸组合也同时属于所有 $\mathcal{C}_\alpha$ 。

多面体是有限个半空间和超平面的交集。半正定锥则可写为无穷个半空间的交集：

$$\mathbb{S}_+^n = \bigcap_{z \neq 0} \{X \in \mathbb{S}^n \mid z^\top X z \geq 0\}.$$



Boyd & Vandenberghe, Section 2.3.1, Example 2.7

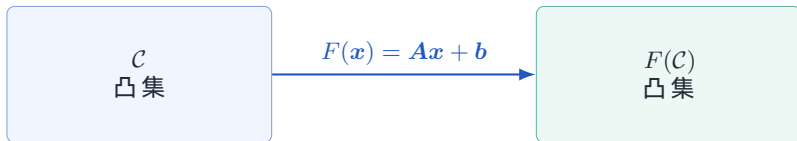
仿射像

设 $F(x) = Ax + b$ 是仿射映射。若 C 是凸集，则其像

$$F(C) = \{Ax + b \mid x \in C\}$$

也是凸集，因为

$$F(\theta x + (1 - \theta)y) = \theta F(x) + (1 - \theta)F(y).$$



$$\alpha C = \{\alpha x \mid x \in C\}, \quad C + a = \{x + a \mid x \in C\}$$

都是该结论的特例。

仿射原像

若 $\mathcal{D} \subseteq \mathbb{R}^m$ 凸，则仿射原像

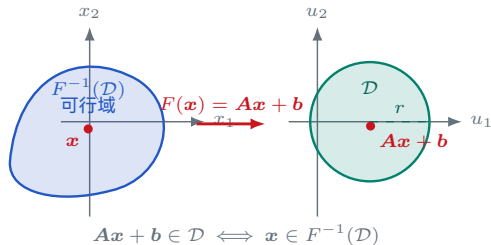
$$F^{-1}(\mathcal{D}) = \{\mathbf{x} \in \mathbb{R}^n \mid \mathbf{Ax} + \mathbf{b} \in \mathcal{D}\}$$

也是凸集。它在约束建模中尤为重要。

例如，范数约束

$$\|\mathbf{Ax} + \mathbf{b}\| \leq r$$

是范数球 $\{\mathbf{u} \mid \|\mathbf{u}\| \leq r\}$ 在映射 $\mathbf{x} \mapsto \mathbf{Ax} + \mathbf{b}$ 下的原像，故为凸集。



积与投影

若 $\mathcal{C}_1, \mathcal{C}_2$ 都凸，则笛卡尔积

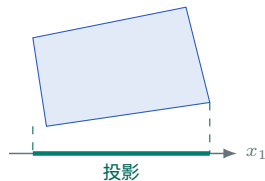
$$\mathcal{C}_1 \times \mathcal{C}_2 = \{(x_1, x_2) \mid x_1 \in \mathcal{C}_1, x_2 \in \mathcal{C}_2\}$$

是凸集。

若 $\mathcal{S} \subseteq \mathbb{R}^m \times \mathbb{R}^n$ 凸，则消去变量后的投影

$$\text{proj}_{x_1} \mathcal{S} = \{x_1 \mid (x_1, x_2) \in \mathcal{S} \text{ for some } x_2\}$$

仍是凸集。



集合和与部分和

两个集合的和 (Minkowski sum) 定义为

$$\mathcal{C}_1 + \mathcal{C}_2 = \{\mathbf{x} + \mathbf{y} \mid \mathbf{x} \in \mathcal{C}_1, \mathbf{y} \in \mathcal{C}_2\}.$$

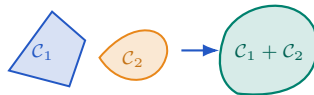
它是 $\mathcal{C}_1 \times \mathcal{C}_2$ 经 $(\mathbf{x}, \mathbf{y}) \mapsto \mathbf{x} + \mathbf{y}$ 的线性像, 所以凸集之和仍为凸集。

更一般地, 若 $\mathcal{S}_i \subseteq \mathbb{R}^n \times \mathbb{R}^m$ 均凸, 则

$$\{(\mathbf{x}, \mathbf{y}_1 + \mathbf{y}_2) \mid (\mathbf{x}, \mathbf{y}_1) \in \mathcal{S}_1, (\mathbf{x}, \mathbf{y}_2) \in \mathcal{S}_2\}$$

称为部分和 (partial sum), 仍为凸集。交集和集合和分别是部分和的两个特殊情形。

集合和



部分和: 固定 $\mathbf{x}$, $\mathbf{y}_1, \mathbf{y}_2$ 相加



线性矩阵不等式

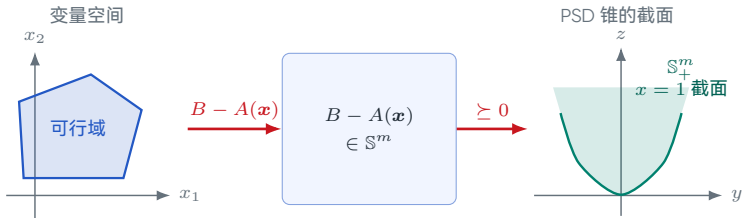
设 $A_i, B \in \mathbb{S}^m$ ，线性矩阵不等式 (linear matrix inequality, LMI)

$$A(\mathbf{x}) = x_1 A_1 + \cdots + x_n A_n \preceq B$$

的可行域是凸集。理由是

$$B - A(\mathbf{x}) \in \mathbb{S}_+^m,$$

即它是 PSD 锥在仿射映射 $\mathbf{x} \mapsto B - A(\mathbf{x})$ 下的原像。



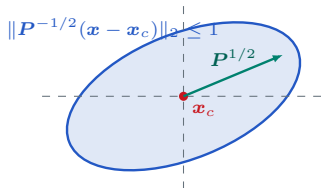
Boyd & Vandenberghe, Section 2.3.2, Example 2.10

两个仿射原像例子

仿射原像把已有凸集直接转化为新可行域。

椭圆是单位球的仿射像，同时也是原像；其原像形式为

$$\|P^{-1/2}(x - x_c)\|_2 \leq 1.$$

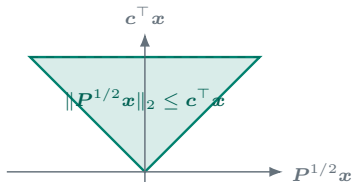


单位球的仿射像

双曲锥 (hyperbolic cone; $P \succeq 0$)

$$\{x \mid x^\top P x \leq (c^\top x)^2, c^\top x \geq 0\}$$

是 SOC 在 $x \mapsto (P^{1/2}x, c^\top x)$ 下的原像。



SOC 的仿射原像

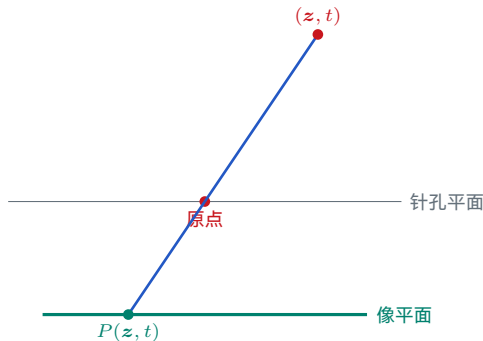
Boyd & Vandenberghe, Section 2.3.2, Examples 2.11–2.12

透视映射

定义在 $t > 0$ 上的**透视映射** (perspective map) 为

$$P(z, t) = \frac{z}{t}, \quad P: \mathbb{R}^n \times \mathbb{R}_{++} \longrightarrow \mathbb{R}^n.$$

它把最后一个坐标归一化为 1。针孔相机给出直观解释：三维点通过原点投影到一个固定像平面，得到的平面坐标正比于 z/t 。



Boyd & Vandenberghe, Section 2.3.3, Remark 2.1

透视保持凸性

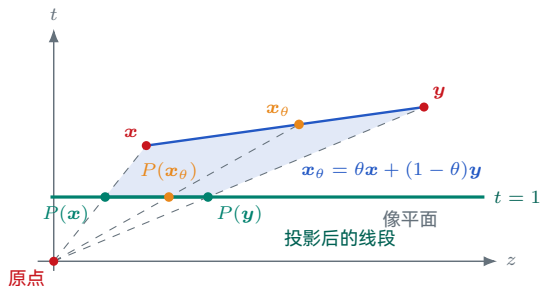
设 $x = (\tilde{x}, x_{n+1})$ 、 $y = (\tilde{y}, y_{n+1})$ ，且最后一维都为正。对 $0 \leq \theta \leq 1$ ，

$$P(\theta x + (1 - \theta)y) = \mu P(x) + (1 - \mu)P(y),$$

其中

$$\mu = \frac{\theta x_{n+1}}{\theta x_{n+1} + (1 - \theta)y_{n+1}} \in [0, 1].$$

因此，透视把一条线段映为一条线段，只是参数 θ 被重新加权为 μ 。若 $C \subseteq \text{dom } P$ 为凸集，则 $P(C)$ 凸；若 D 凸，则 $P^{-1}(D)$ 也凸。

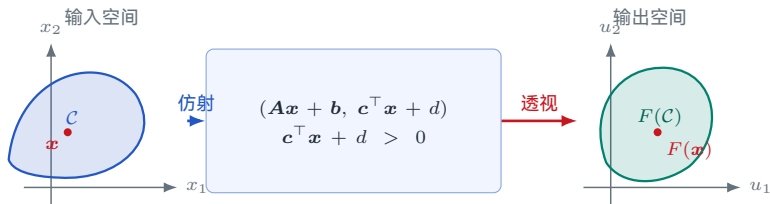


Boyd & Vandenberghe, Section 2.3.3

线性分式映射

线性分式映射 (linear-fractional map) 是仿射映射与透视映射的复合:

$$F(x) = \frac{Ax + b}{c^\top x + d}, \quad \text{dom } F = \{x \mid c^\top x + d > 0\}.$$



若 C 凸且 $C \subseteq \text{dom } F$, 则 $F(C)$ 凸; 凸集在 F 下的原像也凸。

条件概率

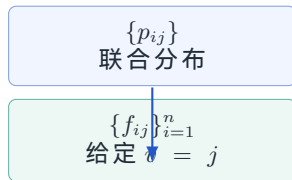
固定一个 j 。令 $p_{ij} = \mathbb{P}(u = i, v = j)$ 是联合概率，且 $\sum_{k=1}^n p_{kj} > 0$ 。则条件分布的第 i 个分量

$$f_{ij} = \mathbb{P}(u = i \mid v = j) = \frac{p_{ij}}{\sum_{k=1}^n p_{kj}}$$

给出向量映射 $p \mapsto (f_{1j}, \dots, f_{nj})$ ，它是关于联合分布 p 的线性分式映射。

因此，若某个不确定性模型允许的联合分布集合 $\mathcal{C}$ 是凸集，并且其上 $\sum_k p_{kj} > 0$ ，那么给定该 j 的条件概率集合也是凸集。

这个例子说明：透视和线性分式映射不是纯粹的几何技巧，它们自然出现于归一化和条件化过程。



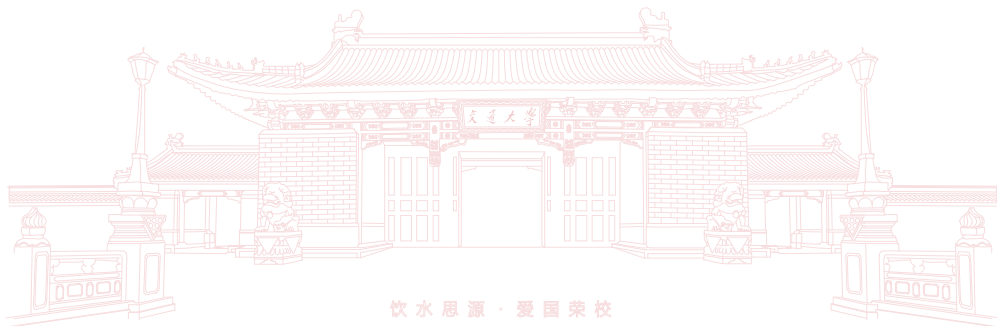
集合演算小结

基本对象	保凸操作
仿射集、半空间、范数球、椭球、多面体、SOC、PSD 锥	交集；仿射像与仿射原像；笛卡尔积；投影；集合和；透视及线性分式映射
建模方式	把可行域拆成基本凸集，并验证所有构造步骤属于上述保凸操作。

集合的表示方式 + 保凸操作 $\implies$ 凸可行域

03

凸函数



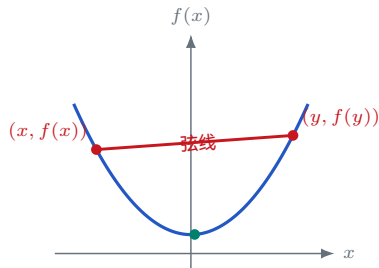
凸函数

设 $\text{dom } f \subseteq \mathbb{R}^n$ 为凸集。函数 f 为**凸函数** (convex function)，若对任意 $x, y \in \text{dom } f$ 及 $0 \leq \theta \leq 1$,

$$f(\theta x + (1 - \theta)y) \leq \theta f(x) + (1 - \theta)f(y).$$

若对 $x \neq y$ 与 $0 < \theta < 1$ 不等式严格成立，则 f 为**严格凸** (strictly convex)。

函数图像上两点的弦线位于图像上方；严格凸时，除端点外弦线严格高于图像。



凹性与直线限制

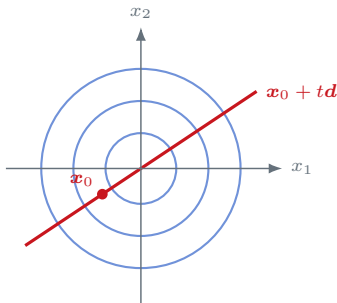
函数 f 为**凹函数** (concave), 若 $-f$ 凸; 严格凹同理。仿射函数 $a^\top x + b$ 同时是凸函数和凹函数。

$$f \text{ 凸} \iff g(t) = f(x_0 + td) \text{ 在其定义域上凸, } \forall x_0, d \neq 0.$$

多元凸性的判断可沿任意直线化为一维问题。若 $\text{dom } f$ 凸,

$$I = \{t \mid x_0 + td \in \text{dom } f\}$$

是区间。



扩展实值函数

若函数 f 只在 $\mathcal{D} \subseteq \mathbb{R}^n$ 上定义，可将它延拓为

$$\tilde{f}(x) = \begin{cases} f(x), & x \in \mathcal{D}, \\ +\infty, & x \notin \mathcal{D}. \end{cases} \quad \text{dom } \tilde{f} = \{x \mid \tilde{f}(x) < +\infty\} = \mathcal{D}.$$

此后可把函数视为定义在整个 $\mathbb{R}^n$ 上，并用扩展实数运算自动记录定义域：

$$\text{dom}(f_1 + f_2) = \text{dom } f_1 \cap \text{dom } f_2.$$

$+\infty$ 表示该点被排除，并不是数值计算中的“大数”。

凸函数的扩展实值形式仍以同一符号 f 记载；这使约束、和式和极小化可以在同一空间中书写。

示性函数

集合 $C \subseteq \mathbb{R}^n$ 的**示性函数** (indicator function) 为

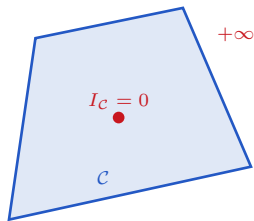
$$I_C(x) = \begin{cases} 0, & x \in C, \\ +\infty, & x \notin C. \end{cases}$$

若 C 是凸集, 则 I_C 是凸函数。

受约束极小化可改写为无显式约束的扩展值问题:

$$\min_{x \in C} f(x) = \min_{x \in \mathbb{R}^n} (f(x) + I_C(x)).$$

这一等式是后续近端算法和对偶理论的基本记号。



一阶条件

设 f 在开凸集 $\mathcal{D}$ 上可微，则

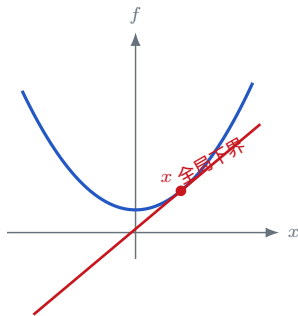
$$f \text{ 凸} \iff f(\mathbf{y}) \geq f(\mathbf{x}) + \nabla f(\mathbf{x})^\top (\mathbf{y} - \mathbf{x}), \quad \forall \mathbf{x}, \mathbf{y} \in \mathcal{D}.$$

右侧是 f 在 x 处的一阶 Taylor 近似。对凸函数，它不只是局部近似，而是 f 的全局仿射下界。

特别地，若 $\nabla f(\mathbf{x}^*) = \mathbf{0}$ ，则

$$f(\mathbf{y}) \geq f(\mathbf{x}^*), \quad \forall \mathbf{y} \in \mathcal{D},$$

故 $\mathbf{x}^*$ 是全局最小点。



尖点处的支撑斜率

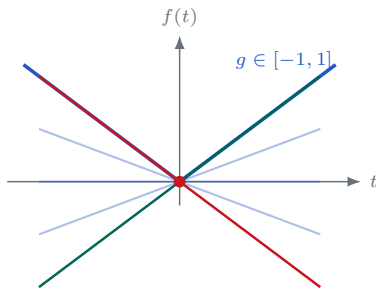
凸函数 $f(t) = |t|$ 在 $t = 0$ 不可微。令 $\ell_g(t) = gt$ ，要求它在全域内位于 f 的下方：

$$|t| \geq gt, \quad \forall t \in \mathbb{R}.$$

当 $t > 0$ 时, $t \geq gt$, 故 $g \leq 1$; 当 $t < 0$ 时, $-t \geq gt$, 故 $g \geq -1$ 。于是

$$-1 \leq g \leq 1.$$

不可微点仍有一族全局支撑斜率。



次梯度拓展；参见 Boyd & Vandenberghe, Section 3.1.3

次梯度

对凸函数 f 与 $x \in \text{dom } f$, 向量 g 是 x 处的 **次梯度** (subgradient), 若

$$f(y) \geq f(x) + g^\top(y - x), \quad \forall y \in \text{dom } f.$$

全部次梯度构成**次微分** (subdifferential) $\partial f(x)$ 。

$$\partial|t| = \begin{cases} \{-1\}, & t < 0, \\ [-1, 1], & t = 0, \\ \{1\}, & t > 0. \end{cases} \quad \partial f(x) = \{\nabla f(x)\} \quad \text{若 } f \text{ 可微}.$$

对于无约束闭凸极小化, x^* 最优当且仅当 $0 \in \partial f(x^*)$ 。

二阶条件

设 $f \in C^2(\mathcal{D})$, 其中 $\mathcal{D}$ 为开凸集。沿方向 d 的限制函数

$$g(t) = f(x + td)$$

满足

$$g''(t) = d^\top \nabla^2 f(x + td) d.$$

因此

$$f \text{ 凸} \iff \nabla^2 f(x) \succeq 0, \quad \forall x \in \mathcal{D}.$$

$\nabla^2 f(x) \succ 0$ 处处成立是严格凸的充分条件;
 $f(x) = x^4$ 表明它不是必要条件。

定义域凸性不能省略: $f(x) = 1/x^2$ 在 $\mathbb{R} \setminus \{0\}$ 上满足 $f''(x) > 0$, 但该函数不是凸函数。

二次函数

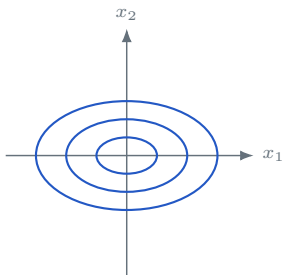
设 $P = P^\top \in \mathbb{S}^n$,

$$f(x) = \frac{1}{2}x^\top Px + q^\top x + r, \quad \nabla f(x) = Px + q, \quad \nabla^2 f(x) = P.$$

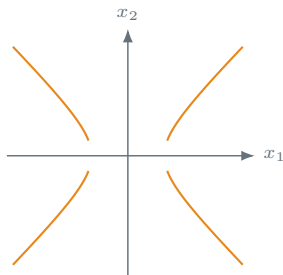
故

$$P \succeq 0 \iff f \text{ 凸}, \quad P \succ 0 \iff f \text{ 严格凸}.$$

$P \succ 0$: 椭圆水平集



P 不定: 双曲线水平集



Boyd & Vandenberghe, Section 3.1.4

基本标量函数

函数	定义域	凸性
e^{ax}	$\mathbb{R}$	凸
x^a	$\mathbb{R}_{++}$	$a \geq 1$ 或 $a \leq 0$ 时凸; $0 \leq a \leq 1$ 时凹
$ x ^p$	$\mathbb{R}$	$p \geq 1$ 时凸
$\log x$	$\mathbb{R}_{++}$	凹
$x \log x$	$\mathbb{R}_+$ ($0 \log 0 = 0$)	凸

$$(x \log x)'' = \frac{1}{x} > 0 \quad (x > 0).$$

函数名称不能替代定义域；同一代数表达式在不同定义域上可以具有不同的凸性结论。

范数与最大值

每个范数都是凸函数。对 $0 \leq \theta \leq 1$,

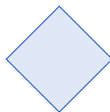
$$\|\theta \mathbf{x} + (1 - \theta) \mathbf{y}\| \leq \theta \|\mathbf{x}\| + (1 - \theta) \|\mathbf{y}\|.$$

分量最大值也是凸函数:

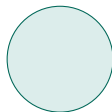
$$\max_i \{\theta x_i + (1 - \theta) y_i\} \leq \theta \max_i x_i + (1 - \theta) \max_i y_i.$$

$$\|\mathbf{x}\|_1, \quad \|\mathbf{x}\|_2, \quad \|\mathbf{x}\|_\infty$$

分别产生菱形、圆形和正方形单位球。



ℓ_1



ℓ_2



ℓ_∞

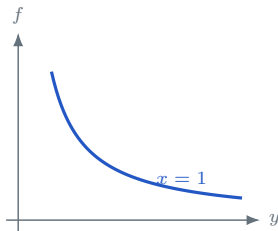
平方除以线性项

$$f(x, y) = \frac{x^2}{y}, \quad \text{dom } f = \mathbb{R} \times \mathbb{R}_{++},$$

是凸函数。其 Hessian 为

$$\nabla^2 f(x, y) = \frac{2}{y^3} \begin{pmatrix} y \\ -x \end{pmatrix} \begin{pmatrix} y \\ -x \end{pmatrix}^\top \succeq 0.$$

此函数也可写为 $x^2/y = y(x/y)^2$ ，是平方函数的透视 (perspective)。它是二阶锥约束、加权最小二乘和后续鲁棒优化中的基本构件。



log-sum-exp

$$f(\mathbf{x}) = \log \left(\sum_{i=1}^n e^{x_i} \right)$$

在 $\mathbb{R}^n$ 上凸, 并满足

$$\max_i x_i \leq f(\mathbf{x}) \leq \max_i x_i + \log n.$$

因此它是 $\max_i x_i$ 的可微近似。

令

$$p_i = \frac{e^{x_i}}{\sum_j e^{x_j}}, \quad \mathbf{p} \succeq \mathbf{0}, \quad \mathbf{1}^\top \mathbf{p} = 1.$$

则

$$\nabla f(\mathbf{x}) = \mathbf{p}, \quad \nabla^2 f(\mathbf{x}) = \text{diag}(\mathbf{p}) - \mathbf{p}\mathbf{p}^\top \succeq 0.$$

softmax 将分数 $\mathbf{x}$ 变为概率 $\mathbf{p}$; log-sum-exp 是机器学习、统计物理和鲁棒建模中反复出现的凸函数。

若干重要函数

几何平均

$$G(\mathbf{x}) = \left(\prod_{i=1}^n x_i \right)^{1/n}, \quad \mathbf{x} \in \mathbb{R}_+^n$$

为凹函数。它刻画乘性效益与配比均衡。

对数行列式

$$f(X) = \log \det X, \quad X \in \mathbb{S}_{++}^n$$

为凹函数。它与体积、信息矩阵和协方差设计有关。

负熵

$$h(\mathbf{x}) = \sum_i x_i \log x_i, \\ \mathbf{x} \succeq 0, \quad 0 \log 0 = 0.$$

为凸函数。其相反数在概率单纯形上为熵。

凸目标的极小化 与 凹目标的极大化

具有相同的凸优化结构。

下水平集

函数 f 的 α -下水平集为

$$C_\alpha = \{x \in \text{dom } f \mid f(x) \leq \alpha\}.$$

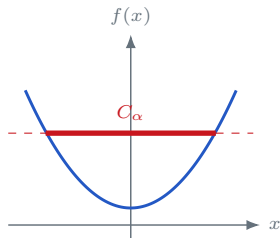
若 f 凸，则每个 C_α 都是凸集；若 f 凹，则其上水平集 $\{x \mid f(x) \geq \alpha\}$ 是凸集。

该性质常用于证明约束集凸性，例如

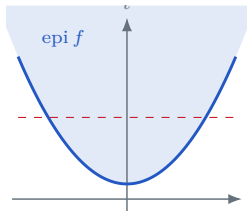
$$\{x \mid \|Ax - b\|_2 \leq r\}$$

是凸集，因为它是凸函数的下水平集。

逆命题不成立： $-e^x$ 的全部下水平集均为区间，但 $-e^x$ 不是凸函数。



上图集与下水平集



上图集:

$$\text{epi } f = \{(\mathbf{x}, t) \mid f(\mathbf{x}) \leq t\}.$$

$$f \text{ 凸} \iff \text{epi } f \text{ 为凸集}.$$

下水平集:

$$C_\alpha = \{\mathbf{x} \mid f(\mathbf{x}) \leq \alpha\}.$$

凸函数的下水平集凸；凹函数的上水平集凸。

Jensen 不等式

若 f 凸, $\theta_i \geq 0$, 且 $\sum_{i=1}^k \theta_i = 1$, 则

$$f\left(\sum_{i=1}^k \theta_i x_i\right) \leq \sum_{i=1}^k \theta_i f(x_i).$$

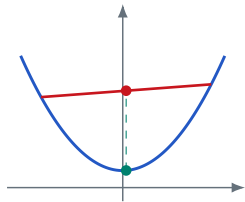
若 X 是取值在 $\text{dom } f$ 中的随机变量, 则

$$f(\mathbb{E}X) \leq \mathbb{E}f(X).$$

对零均值随机扰动 Z ,

$$\mathbb{E}f(x + Z) \geq f(x).$$

因此, 随机化不会在平均意义下降低凸代价。对凹函数, 所有不等号反向。



保凸运算

从已知凸函数构造新凸函数的主要运算包括：

运算	保凸条件
非负加权和	f_i 凸, $w_i \geq 0$
仿射复合	f 凸, $g(\mathbf{x}) = f(\mathbf{Ax} + \mathbf{b})$
逐点上确界	$f(\mathbf{x}) = \sup_{\alpha} f_{\alpha}(\mathbf{x})$
函数复合	外函数的单调性与内函数的凸/凹性匹配
部分最小化	$f(\mathbf{x}, \mathbf{y})$ 联合凸, $\mathbf{y}$ 的可行集凸
透视	$g(\mathbf{x}, t) = tf(\mathbf{x}/t)$, $t > 0$

这些规则把有限个基本函数组合为工程模型的目标与约束。

非负加权和

若 f_i 均为凸函数, 且 $w_i \geq 0$, 则

$$f(\mathbf{x}) = \sum_{i=1}^m w_i f_i(\mathbf{x})$$

为凸函数。若至少一个正权重对应严格凸函数, 则相应条件下可得到严格凸性。

加权最小二乘

$$\sum_{i=1}^m w_i (\mathbf{a}_i^\top \mathbf{x} - b_i)^2, \quad w_i \geq 0,$$

由平方仿射函数的非负加权和构成, 因此凸。

在连续情形, 若 $f(x, y)$ 对 x 凸且 $w(y) \geq 0$, 则

$$\int_A w(y) f(x, y) dy$$

仍对 x 凸。

仿射复合

若 $f: \mathbb{R}^m \rightarrow \mathbb{R}$ 凸, $A \in \mathbb{R}^{m \times n}$, 则

$$g(x) = f(Ax + b), \quad \text{dom } g = \{x \mid Ax + b \in \text{dom } f\},$$

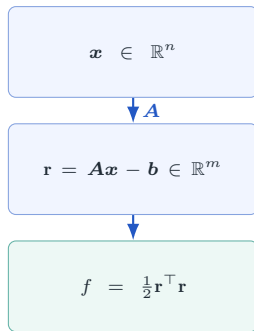
也是凸函数。

传感器标定可写为

$$\min_x \frac{1}{2} \|Ax - b\|_2^2, \quad A \in \mathbb{R}^{m \times n}, x \in \mathbb{R}^n, b \in \mathbb{R}^m.$$

其梯度为

$$\nabla f(x) = A^\top (Ax - b).$$



逐点最大值

若 $f_1, \dots, f_k$ 均为凸函数，则

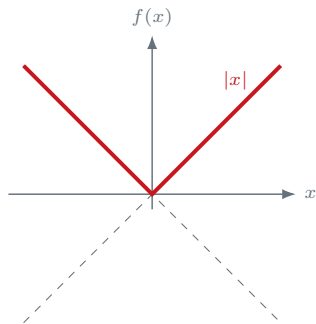
$$f(x) = \max_{i=1, \dots, k} f_i(x)$$

也是凸函数。更一般地，任意凸函数族的逐点上确界仍凸。

$$|x| = \max\{x, -x\}, \quad f(x) = \max_i (\mathbf{a}_i^\top \mathbf{x} + b_i).$$

后一形式给出分段仿射凸函数。上确界规则还产生支持函数

$$S_C(\mathbf{y}) = \sup_{\mathbf{x} \in C} \mathbf{y}^\top \mathbf{x}.$$



标量复合

令 $f = h \circ g$ 。下列四条规则在一般定义域下仍成立，其中单调性指扩展值函数的单调性：

$$\begin{aligned}h \text{ 凸、非减}, g \text{ 凸} &\Rightarrow h \circ g \text{ 凸}, \\h \text{ 凸、非增}, g \text{ 凹} &\Rightarrow h \circ g \text{ 凸}, \\h \text{ 凹、非减}, g \text{ 凹} &\Rightarrow h \circ g \text{ 凹}, \\h \text{ 凹、非增}, g \text{ 凸} &\Rightarrow h \circ g \text{ 凹}.\end{aligned}$$

若 h, g 二次可微，则

$$(h \circ g)'' = h''(g)(g')^2 + h'(g)g''.$$

该式解释了凸性、凹性与单调性必须同时匹配。

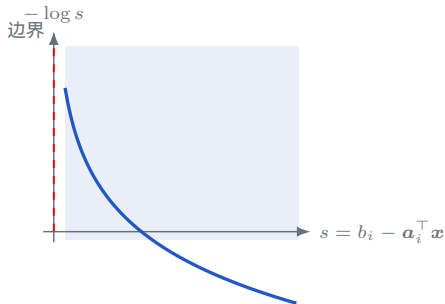
复合规则的例子

$$g \text{ convex} \implies e^{g(\mathbf{x})} \text{ convex}$$

$$g \text{ concave}, g > 0 \implies \log g(\mathbf{x}) \text{ concave}$$

$$g \text{ concave}, g > 0 \implies 1/g(\mathbf{x}) \text{ convex}$$

$$g \text{ convex} \implies -\log(-g(\mathbf{x})) \text{ convex on } g < 0$$



$$\phi(\mathbf{x}) = -\sum_{i=1}^m \log(b_i - \mathbf{a}_i^\top \mathbf{x})$$

其定义域为

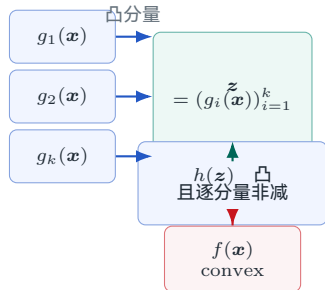
$$\{\mathbf{x} \mid \mathbf{a}_i^\top \mathbf{x} < b_i, i = 1, \dots, m\}$$

且在该定义域上为凸函数。它是线性不等式的 barrier function; 接近边界时函数值发散。

Boyd & Vandenberghe, Sections 3.2.4–3.2.5

向量复合

设 $f(\mathbf{x}) = h(g_1(\mathbf{x}), \dots, g_k(\mathbf{x}))$ 。若 h 为凸函数且在每个分量上非减，并且所有 g_i 为凸函数，则 f 为凸函数。



两个常用的外层函数

例如 $h(\mathbf{z}) = \log \sum_i e^{z_i}$ 为凸函数且逐分量非减，故

$$\log \left(\sum_{i=1}^k e^{g_i(\mathbf{x})} \right)$$

在每个 g_i 为凸函数时仍为凸函数。

若 $p \geq 1$ ，每个 $g_i \geq 0$ 且为凸函数，则

$$\left(\sum_{i=1}^k g_i(\mathbf{x})^p \right)^{1/p}$$

也为凸函数。它是由逐分量非减的 p -范数与凸分量复合得到的。

部分最小化

若 $f(x, y)$ 关于 (x, y) 联合凸, C 凸且非空, 且 $g(x) > -\infty$ 对所有 x 成立, 则

$$g(x) = \inf_{y \in C} f(x, y)$$

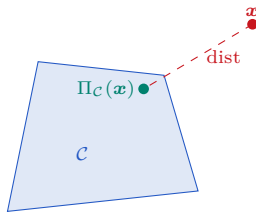
为凸函数, 且

$$\text{dom } g = \{x \mid \exists y \in C, (x, y) \in \text{dom } f\}.$$

取凸集 C , 则到该集合的距离

$$\text{dist}(x, C) = \inf_{y \in C} \|x - y\|$$

是凸函数: $\|x - y\|$ 对 (x, y) 联合凸。



透视函数

函数 $f: \mathbb{R}^n \rightarrow \mathbb{R}$ 的**透视** (perspective) 定义为

$$g(\mathbf{x}, t) = tf(\mathbf{x}/t), \quad \text{dom } g = \{(\mathbf{x}, t) \mid t > 0, \mathbf{x}/t \in \text{dom } f\}.$$

若 f 凸, 则 g 凸。

对 $f(\mathbf{x}) = \|\mathbf{x}\|_2^2$,

$$g(\mathbf{x}, t) = \frac{\|\mathbf{x}\|_2^2}{t}, \quad t > 0,$$

即前述 quadratic-over-linear 函数。

对 $f(x) = -\log x$, 透视给出

$$-t \log(x/t) = t \log(t/x),$$

即相对熵的基本形式。



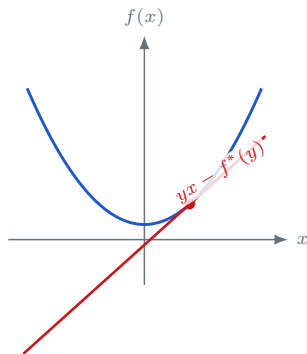
共轭函数

扩展实值函数 $f: \mathbb{R}^n \rightarrow (-\infty, +\infty]$ 的 Fenchel 共轭为

$$f^*(y) = \sup_{x \in \mathbb{R}^n} (y^\top x - f(x)).$$

对固定 y ，它考察斜率为 y 的线性函数与 f 的最大差距。 $\text{dom } f^*$ 是上述上确界有限的 y 所组成的集合。

共轭将函数的“自变量”解释为一个斜率或价格向量。



Boyd & Vandenberghe, Section 3.3.1

共轭为何凸

对每个固定的 x ，函数

$$\mathbf{y} \mapsto \mathbf{y}^\top \mathbf{x} - f(x)$$

是 $\mathbf{y}$ 的仿射函数。故

$$f^*(\mathbf{y}) = \sup_x \{\mathbf{y}^\top \mathbf{x} - f(x)\}$$

是仿射函数族的逐点上确界，必为凸函数。

这一定义不要求 f 本身凸。若 f 已用扩展实值形式书写，则 $x \notin \text{dom } f$ 时被减去的 $+\infty$ 自动排除该点。

f^* 收集 f 的全部仿射下界。后续对偶理论正是利用这些下界构造原问题最优值的下界。

二次函数的共轭

令

$$f(x) = \frac{1}{2}x^\top Qx, \quad Q \succ 0.$$

对给定 y , 最大化

$$y^\top x - \frac{1}{2}x^\top Qx.$$

一阶条件给出

$$y - Qx = 0, \quad x^* = Q^{-1}y.$$

代回得到

$$f^*(y) = \frac{1}{2}y^\top Q^{-1}y.$$

若 $f(x) = \frac{1}{2}x^\top Qx + q^\top x + r$, 则平移和常数项由共轭的仿射变换规则处理; 本页的二次型计算是最常用的基本情形。

指数与负熵

对 $f(x) = e^x$, 令 $yx - e^x$ 对 x 的导数为零:

$$y - e^x = 0.$$

当 $y > 0$ 时最大点为 $x = \log y$, 并且

$$f^*(y) = y \log y - y.$$

结合边界情形,

$$(e^x)^*(y) = \begin{cases} y \log y - y, & y \geq 0, \\ +\infty, & y < 0, \end{cases}$$

其中 $0 \log 0 = 0$ 。指数函数的共轭是带线性平移的负熵; 这一对偶关系解释了 softmax、最大熵建模和 log-sum-exp 的紧密联系。

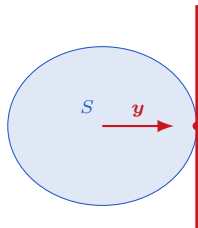
示性函数与支持函数

对任意非空集合 $S \subseteq \mathbb{R}^n$ ，其示性函数满足

$$I_S^*(\mathbf{y}) = \sup_{\mathbf{x} \in S} \mathbf{y}^\top \mathbf{x} = S_S(\mathbf{y}),$$

右侧称为 S 的**支持函数** (support function)。

因此，共轭将“集合约束”转化为“沿方向 $\mathbf{y}$ 的最大线性收益”。这连接了前面的半空间、支持超平面与后续鲁棒约束。



范数的共轭

令 $\|\cdot\|_*$ 为 $\|\cdot\|$ 的对偶范数:

$$\|y\|_* = \sup_{\|x\| \leq 1} y^\top x.$$

则

$$(\|x\|)^*(y) = \begin{cases} 0, & \|y\|_* \leq 1, \\ +\infty, & \|y\|_* > 1. \end{cases}$$

原因是 Hölder 不等式给出

$$y^\top x - \|x\| \leq (\|y\|_* - 1)\|x\|.$$

若 $\|y\|_* \leq 1$, 最大值在 $x = 0$ 取得。

对二范数,

$$\left(\frac{1}{2}\|x\|_2^2\right)^* = \frac{1}{2}\|y\|_2^2.$$

范数正则化与对偶球约束因而是一对共轭结构。

Fenchel–Young 不等式

由共轭定义可直接得到，对任意 x, y ,

$$f^*(y) \geq y^\top x - f(x),$$

即

$$f(x) + f^*(y) \geq x^\top y.$$

对于 $f(x) = \frac{1}{2}x^\top Qx$, $Q \succ 0$,

$$x^\top y \leq \frac{1}{2}x^\top Qx + \frac{1}{2}y^\top Q^{-1}y.$$

它是加权 Cauchy–Schwarz 不等式的一种形式。

等号表示仿射下界 $z \mapsto y^\top z - f^*(y)$ 在 x 处接触 f 。对适当、闭、凸且在 x 可微的 f , 等号当且仅当 $y = \nabla f(x)$ 。

双共轭与梯度

若 f 是适当、闭、凸函数，则

$$f^{**} = f.$$

闭性在这里指 $\text{epi } f$ 是闭集。双共轭说明：一个闭凸函数可由其全部仿射下界精确恢复。

若 f 可微且 $\text{dom } f = \mathbb{R}^n$ ，并且上确界在 x^* 处达到，则

$$y = \nabla f(x^*), \quad f^*(y) = x^{*\top} \nabla f(x^*) - f(x^*).$$

更一般地，对适当、闭、凸函数，

$$y \in \partial f(x) \iff f(x) + f^*(y) = x^\top y.$$

这将本章的次微分与共轭函数连接起来。

共轭的运算规则

设 $a > 0$ 。若 $g(x) = af(x) + b$, 则

$$g^*(y) = af^*(y/a) - b.$$

若 A 非奇异, $g(x) = f(Ax + b)$, 则

$$g^*(y) = f^*(A^{-\top}y) - b^{\top}A^{-\top}y.$$

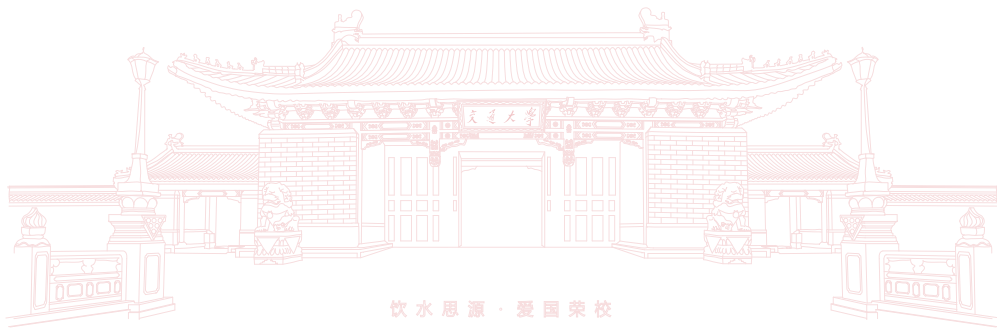
若 $f(u, v) = f_1(u) + f_2(v)$, 则

$$f^*(w, z) = f_1^*(w) + f_2^*(z).$$

至此, 凸函数可以由定义、导数、上图集和保凸运算加以验证; 共轭函数则给出它们的仿射下界表示。这些结构将在下一章的凸优化问题与随后对偶理论中同时出现。

04

凸优化问题



饮水思源 · 爱国荣校

优化问题

$$\begin{array}{ll}\underset{\boldsymbol{x} \in \mathbb{R}^n}{\text{minimize}} & f_0(\boldsymbol{x}) \\ \text{subject to} & f_i(\boldsymbol{x}) \leq 0, \quad i = 1, \dots, m, \\ & h_j(\boldsymbol{x}) = 0, \quad j = 1, \dots, p.\end{array}$$

- ▣ $\boldsymbol{x}$: optimization variable (决策变量);
- ▣ f_0 : objective/cost function (目标/代价函数);
- ▣ $f_i \leq 0$ 、 $h_j = 0$: inequality / equality constraints。

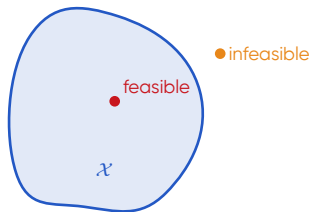
问题由函数及其数据给出；求解的对象是满足约束的变量值，而不是一个孤立的公式。

定义域与可行性

问题的定义域是所有函数同时有定义的点：

$$D = \bigcap_{i=0}^m \text{dom } f_i \cap \bigcap_{j=1}^p \text{dom } h_j.$$

- ▣ $x \in D$ 且满足全部约束，称为 feasible point;
- ▣ 所有 feasible points 构成 feasible set $\mathcal{X}$;
- ▣ $\mathcal{X} \neq \emptyset$ 时问题 feasible，否则 infeasible。



最优值与三种状态

$$p^* = \inf\{f_0(x) \mid x \in \mathcal{X}\}.$$

$$\begin{aligned}\mathcal{X} &= \emptyset \\ \text{infeasible} \\ p^* &= +\infty\end{aligned}$$

$$\begin{aligned}f_0(x_k) &\rightarrow -\infty \\ \text{unbounded below} \\ p^* &= -\infty\end{aligned}$$

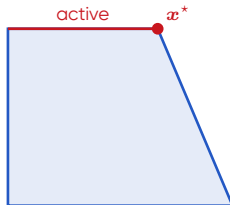
$$\begin{aligned}p^* &\in \mathbb{R} \\ \text{可能取到, 也可能不取到}\end{aligned}$$

例如, 在 $x > 0$ 上: $\min 1/x$ 的最优值为 0 但不取到; $\min(-\log x)$ 无下界; $\min x \log x$ 的最优点为 $1/e$ 。

最优点与约束状态

$$\mathcal{X}_{\text{opt}} = \{x \in \mathcal{X} \mid f_0(x) = p^*\}.$$

- ▣ 若 $\mathcal{X}_{\text{opt}} \neq \emptyset$, 称最优值 attained;
- ▣ $f_0(x) \leq p^* + \epsilon$ 的 feasible x 称为 ϵ -suboptimal;
- ▣ 若 $f_i(x) = 0$, 第 i 个不等式在 x 处 active; 若 $f_i(x) < 0$, 则 inactive;
- ▣ 删除后不改变可行域的约束称 redundant。



可行性问题

当 $f_0 \equiv 0$ 时，优化问题退化为

$$\begin{array}{ll}\text{find} & \boldsymbol{x} \\ \text{subject to} & f_i(\boldsymbol{x}) \leq 0, \quad h_j(\boldsymbol{x}) = 0.\end{array}$$

此时只问约束是否一致，以及在一致时给出一个可行点。

许多优化算法的一个基本步骤就是反复求解可行性问题；
拟凸优化也可化为一系列可行性检验。

$\mathcal{X} \neq \emptyset$: 任取 $\boldsymbol{x} \in \mathcal{X}$ 即为解;
 $\mathcal{X} = \emptyset$: 没有可行解。

标准形式

标准形式将不等式和等式的右端统一写为零：

$$f_i(\mathbf{x}) \leq 0, \quad h_j(\mathbf{x}) = 0.$$

箱约束

$$\ell_i \leq x_i \leq u_i$$

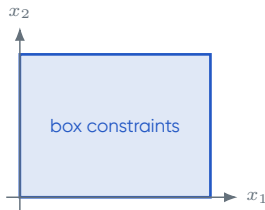
等价于两组不等式

$$\ell_i - x_i \leq 0, \quad x_i - u_i \leq 0.$$

最大化问题

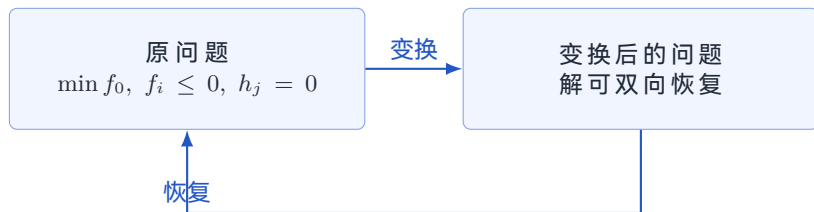
$$\max f_0(\mathbf{x})$$

通过最小化 $-f_0(\mathbf{x})$ 处理。



等价问题

若能从一个问题的解直接恢复另一个问题的解，反之亦然，则称二者 equivalent。



常用变换包括：正比例缩放、单调函数变换、变量替换、引入 slack variables、消去或引入等式约束。

不改变解的变换

对于 $\alpha_i > 0, \beta_j \neq 0$,

$$\min \alpha_0 f_0(\mathbf{x}), \quad \alpha_i f_i(\mathbf{x}) \leq 0, \quad \beta_j h_j(\mathbf{x}) = 0$$

与原问题有相同的可行域和最优点。

更一般地，只要 ψ_0 严格单调递增，且

$$\psi_i(u) \leq 0 \iff u \leq 0,$$

就可用 $\psi_i \circ f_i$ 代替 f_i 。

$$\arg \min_{\mathbf{x}} \|\mathbf{Ax} - \mathbf{b}\|_2 = \arg \min_{\mathbf{x}} \|\mathbf{Ax} - \mathbf{b}\|_2^2.$$

最优点相同；后者是可微二次函数。

变量替换与消去等式

若 $\phi: \tilde{\mathcal{D}} \rightarrow \mathcal{D}$ 是双射, 令 $x = \phi(z)$, 则以

$$\tilde{f}_i(z) = f_i(\phi(z)), \quad \tilde{h}_j(z) = h_j(\phi(z))$$

构成的 z -问题与原问题等价。

特别地, 若 $Ax = b$ 可行, 取一个特解 x_0 和 $\text{Range}(F) = \text{Null}(A)$, 则

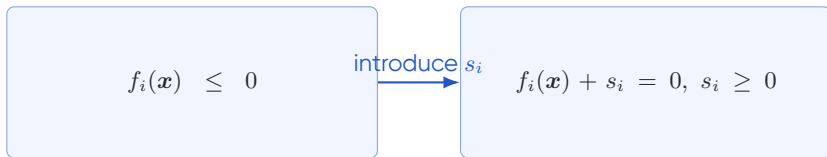
$$x = Fz + x_0$$

参数化全部可行点; 代入后消去线性等式。

Slack 与辅助变量

不等式 $f_i(x) \leq 0$ 等价于存在 $s_i \geq 0$, 使

$$f_i(x) + s_i = 0.$$



若出现复合项 $f_i(A_i x + b_i)$, 也可引入

$$y_i = A_i x + b_i$$

将其写为 $f_i(y_i)$ 加上线性等式。这种拆分在建模和算法中很常用。

参数模型与 Oracle

Parameter description

显式给定问题数据。例如

$$f_0(x) = \frac{1}{2}x^T Px + q^T x + r,$$

数据为 P, q, r 。

$P \quad q \quad r$
问题数据

Oracle description

不显式给出公式；输入 x 后，子程序返回 $f(x)$ ，通常也返回 $\nabla f(x)$ 或次梯度。



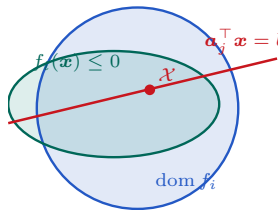
后续算法可只依赖 oracle；利用特定参数结构则往往能得到更高效率。

凸优化标准形式

$$\begin{array}{ll}\underset{\mathbf{x}}{\text{minimize}} & f_0(\mathbf{x}) \\ \text{subject to} & f_i(\mathbf{x}) \leq 0, \quad i = 1, \dots, m, \\ & \mathbf{a}_j^\top \mathbf{x} = b_j, \quad j = 1, \dots, p.\end{array}$$

这是 convex optimization problem, 当且仅当

$f_0, f_1, \dots, f_m$ 均为凸函数, 且等式约束为仿射函数。



可行域是各凸函数定义域、凸下水平集和仿射超平面的交集, 因而是凸集。

解集的结构

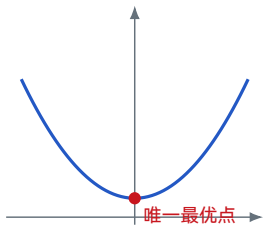
若 f_0 为凸函数，则任意 ϵ -suboptimal set 都是凸集；尤其，

$$\mathcal{X}_{\text{opt}} = \{x \in \mathcal{X} \mid f_0(x) = p^*\}$$

是凸集。

若 f_0 严格凸，则 $\mathcal{X}_{\text{opt}}$ 至多含一个点。

最大化凹函数等价于最小化其相反数，因此 concave maximization 也属于凸优化的范围。



表示方式同样重要

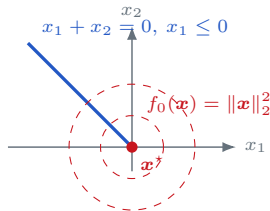
“凸函数在凸集上最小化”还不足以成为教材中的标准凸优化形式：约束必须以 convex inequalities 和 affine equalities 表示。

$$\begin{aligned} \min_{\mathbf{x} \in \mathbb{R}^2} \quad & x_1^2 + x_2^2 \\ \text{subject to} \quad & \frac{x_1}{1+x_2^2} \leq 0, \quad (x_1+x_2)^2 = 0 \end{aligned}$$

可行域本身凸，却不是标准凸形式。它等价于

$$x_1 \leq 0, \quad x_1 + x_2 = 0,$$

后者才是凸优化问题的有效表示。



局部最优即全局最优

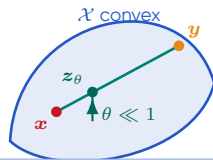
设 x 是局部最优。若有 $y \in \mathcal{X}$ 满足 $f_0(y) < f_0(x)$, 令

$$z_\theta = (1 - \theta)x + \theta y, \quad 0 < \theta \ll 1.$$

由可行域凸性, z_θ 可行; 由目标凸性,

$$f_0(z_\theta) \leq (1 - \theta)f_0(x) + \theta f_0(y) < f_0(x).$$

这与局部最优性矛盾。



$f_0(y) < f_0(x)$
会在 x 的邻域内产生更优可行点

一阶最优条件

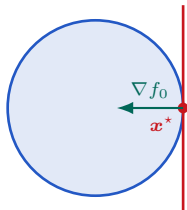
设 f_0 为可微凸函数， $\mathcal{X}$ 为非空凸可行域。则

$$x^* \text{ 最优} \iff \nabla f_0(x^*)^\top (y - x^*) \geq 0, \quad \forall y \in \mathcal{X}.$$

由凸函数一阶下界

$$f_0(y) \geq f_0(x) + \nabla f_0(x)^\top (y - x)$$

可立即验证充分性；必要性来自沿可行线段的方向导数。



无约束凸二次问题

对

$$f(x) = \frac{1}{2}x^\top Px + q^\top x + r, \quad P \succeq 0,$$

最优性的充要条件是

$$Px + q = 0.$$

情形	结论
$q \notin \text{Range}(P)$	无解, 问题无下界
$P \succ 0$	唯一解 $x^* = -P^{-1}q$
P 奇异且 $q \in \text{Range}(P)$	$\mathcal{X}_{\text{opt}} = -P^\dagger q + \text{Null}(P)$

仅有线性等式

$$\min f_0(\mathbf{x}) \quad \text{subject to} \quad \mathbf{Ax} = \mathbf{b}.$$

所有可行方向 $\mathbf{v}$ 属于 $\text{Null}(\mathbf{A})$ 。一阶最优条件为

$$\nabla f_0(\mathbf{x}^*)^\top \mathbf{v} = 0, \quad \forall \mathbf{v} \in \text{Null}(\mathbf{A}).$$

等价地,

$$\nabla f_0(\mathbf{x}^*) \in \text{Range}(\mathbf{A}^\top) \iff \exists \boldsymbol{\nu} : \nabla f_0(\mathbf{x}^*) + \mathbf{A}^\top \boldsymbol{\nu} = \mathbf{0},$$

并同时满足 $\mathbf{Ax}^* = \mathbf{b}$ 。这已是 Lagrange multiplier 条件的基本形态。

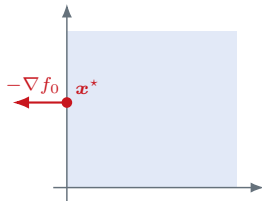
非负正交锥

$$\min_{x \succeq 0} f_0(x).$$

可微凸情形的一阶最优条件等价于

$$x^* \succeq 0, \quad \nabla f_0(x^*) \succeq 0, \quad x_i^* \nabla_i f_0(x^*) = 0 \quad (i = 1, \dots, n).$$

最后一个关系称为 complementarity: 每个坐标上, 变量或梯度至少有一个为零。



等价凸问题

下列操作保持 convexity:

- ▣ 消去 $Ax = b$: 代入 $x = Fz + x_0$, 凸函数与仿射映射复合仍凸;
- ▣ 引入 $y = Ax + b$: 增加仿射等式, 保留凸性;
- ▣ 对 affine inequality 引入 slack variable;
- ▣ 将目标改写为 epigraph form;
- ▣ 对联合凸函数在部分变量上最小化 (partial minimization)。

Epigraph 形式

引入标量 t 作为目标函数的上界。一般凸问题可等价改写为线性目标最小化，附加约束为

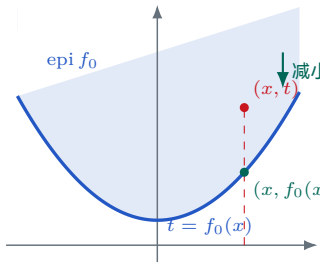
$$f_0(x) - t \leq 0.$$

由于 $f_0(x) - t$ 仍为凸函数，改写后的问题保持凸性。

原问题
 $\min_x f_0(x)$



Epigraph 形式
 $\min_{x,t} t$
s.t. $f_0(x) \leq t$



拟凸优化

在 quasiconvex optimization 中，约束仍为凸不等式和仿射等式，但目标 f_0 仅为 quasiconvex。

$$\min f_0(\mathbf{x}) \quad \text{subject to} \quad f_i(\mathbf{x}) \leq 0, \mathbf{Ax} = \mathbf{b}.$$

- 最优集仍凸；严格局部最优点是全局最优，但一般局部最优点未必全局最优；
- 对可微 f_0 ， $\nabla f_0(\mathbf{x})^\top (\mathbf{y} - \mathbf{x}) > 0$ 是最优的充分条件，不再是充要条件；
- 典型例子是 linear-fractional objective。

拟凸问题的可行性检验

若凸函数族 ϕ_t 满足

$$f_0(\mathbf{x}) \leq t \iff \phi_t(\mathbf{x}) \leq 0,$$

则对于给定 t , 检验

$$\text{find } \mathbf{x} \quad \text{subject to} \quad \phi_t(\mathbf{x}) \leq 0, \quad f_i(\mathbf{x}) \leq 0, \quad \mathbf{Ax} = \mathbf{b}$$

是一个 convex feasibility problem。

随着 t 增大, 可行性单调出现; 因而可以对 t 二分搜索, 逼近最优值 p^* 。

线性规划的三种形式

General form

$$\min c^\top x + d$$

$$Gx \preceq h, Ax = b$$

Standard form

$$\min c^\top x$$

$$Ax = b, x \succeq 0$$

Inequality form

$$\min c^\top x$$

$$Ax \preceq b$$

所有目标和约束函数都是仿射函数；可行域是 polyhedron。

转为 LP 标准形式

从 general form 出发, 先对 $Gx \preceq h$ 引入 slack $\succeq 0$:

$$Gx + s = h.$$

再将自由变量写为

$$x = x^+ - x^-, \quad x^+, x^- \succeq 0.$$

得到

$$\begin{aligned} \min \quad & c^\top x^+ - c^\top x^- + d \\ \text{subject to} \quad & Gx^+ - Gx^- + s = h, \\ & Ax^+ - Ax^- = b, \\ & x^+, x^-, s \succeq 0. \end{aligned}$$

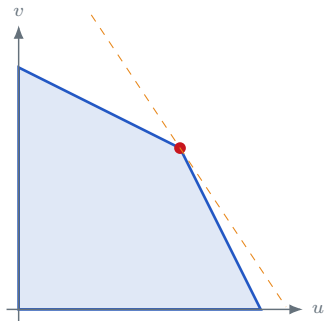
生产计划

设 u, v 为两类产品产量；每批收益为 3, 2，两道工序容量均为 8：

$$\begin{aligned} \min_{u,v} \quad & -3u - 2v \\ \text{subject to} \quad & 2u + v \leq 8, \quad u + 2v \leq 8, \\ & u, v \geq 0. \end{aligned}$$

目标和全部约束均为仿射，故为 LP。最优点处两道容量约束同时 active：

$$(u^*, v^*) = (8/3, 8/3).$$



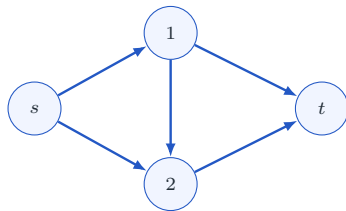
自编示例；参见 Boyd & Vandenberghe, Section 4.3.1

网络流

令 x_e 为边 e 的流量， B 为关联矩阵； d 表示节点净供给。

$$\begin{aligned} \min_x \quad & c^\top x \\ \text{subject to} \quad & Bx = d, \\ & 0 \preceq x \preceq u. \end{aligned}$$

$Bx = d$ 是流量守恒；上下容量可写成线性不等式。道路、物流、通信网络具有相同的 LP 结构。



多面体的 Chebyshev 中心

对多面体

$$\mathcal{P} = \{x \mid a_i^\top x \leq b_i, i = 1, \dots, m\},$$

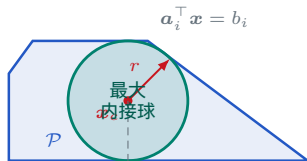
求其内接最大欧氏球的中心 x_c 与半径 r 。

$$\sup_{\|u\|_2 \leq r} a_i^\top (x_c + u) = a_i^\top x_c + r \|a_i\|_2 \leq b_i.$$

因此

$$\max_{x_c, r} r \quad \text{subject to} \quad a_i^\top x_c + r \|a_i\|_2 \leq b_i$$

是 LP。几何距离问题由此变成线性规划。



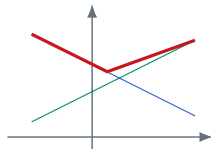
分段线性目标

$$f(\mathbf{x}) = \max_{i=1,\dots,m} (\mathbf{a}_i^\top \mathbf{x} + b_i)$$

是 convex piecewise-linear function。利用 epigraph:

$$\begin{aligned} \min_{\mathbf{x}, t} \quad & t \\ \text{subject to} \quad & \mathbf{a}_i^\top \mathbf{x} + b_i \leq t, \quad i = 1, \dots, m. \end{aligned}$$

每个 affine piece 生成一条线性不等式；最小化最大值因可作为 LP 求解。



线性分式规划

$$\begin{aligned} \min_x \quad & \frac{c^\top x + d}{e^\top x + f} \\ \text{subject to} \quad & Gx \preceq h, \quad Ax = b, \\ & e^\top x + f > 0. \end{aligned}$$

该 objective 是 linear-fractional function, 既 quasiconvex 又 quasiconcave (quasilinear)。

例如, “每单位资源的收益/能耗”或“单位时间的产出”常导致这类比例目标。

分子与分母均为仿射函数;
分母的正性是定义域的一部分。

线性分式到 LP

令

$$\mathbf{y} = \frac{\mathbf{x}}{\mathbf{e}^\top \mathbf{x} + f}, \quad z = \frac{1}{\mathbf{e}^\top \mathbf{x} + f}.$$

则 $\mathbf{x} = \mathbf{y}/z$, 并得到等价 LP:

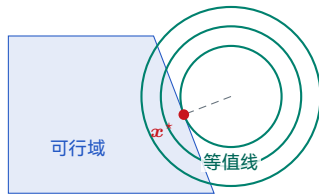
$$\begin{aligned} \min_{\mathbf{y}, z} \quad & \mathbf{c}^\top \mathbf{y} + dz \\ \text{subject to} \quad & \mathbf{G}\mathbf{y} - \mathbf{h}z \preceq \mathbf{0}, \quad \mathbf{A}\mathbf{y} - \mathbf{b}z = \mathbf{0}, \\ & \mathbf{e}^\top \mathbf{y} + fz = 1, \quad z \geq 0. \end{aligned}$$

若 $z > 0$, 以 $\mathbf{x} = \mathbf{y}/z$ 恢复原变量; 若 $z = 0$, 它对应原问题的一条极限方向, 原问题中的可行点可以逼近相同的目标值。因此两个问题的最优值相同。

二次规划

$$\begin{aligned} \min_x \quad & \frac{1}{2} \mathbf{x}^\top \mathbf{P} \mathbf{x} + \mathbf{q}^\top \mathbf{x} + r \\ \text{subject to} \quad & \mathbf{G} \mathbf{x} \preceq \mathbf{h}, \quad \mathbf{A} \mathbf{x} = \mathbf{b}, \quad \mathbf{P} \succeq \mathbf{0}. \end{aligned}$$

QP 是在 polyhedron 上最小化 convex quadratic function。取 $\mathbf{P} = \mathbf{0}$ 即得到 LP。



QCQP 与模型层次

若不等式也为凸二次函数，则

$$\begin{aligned} \min_x \quad & \frac{1}{2} \mathbf{x}^\top \mathbf{P}_0 \mathbf{x} + \mathbf{q}_0^\top \mathbf{x} + r_0 \\ \text{subject to} \quad & \frac{1}{2} \mathbf{x}^\top \mathbf{P}_i \mathbf{x} + \mathbf{q}_i^\top \mathbf{x} + r_i \leq 0, \quad i = 1, \dots, m, \quad \mathbf{P}_i \succeq 0, \\ & \mathbf{A} \mathbf{x} = \mathbf{b}, \end{aligned}$$

称为 quadratically constrained quadratic program (QCQP)。



最小二乘与约束回归

$$\|Ax - b\|_2^2 = x^\top A^\top Ax - 2b^\top Ax + b^\top b.$$

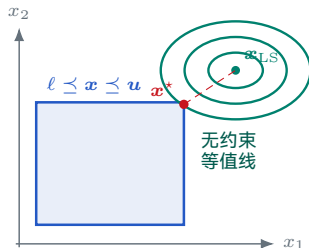
因此 least squares 是无约束 QP; 若 A 满列秩, 解为

$$x^* = (A^\top A)^{-1} A^\top b.$$

加入箱约束后,

$$\min_x \|Ax - b\|_2^2 \quad \text{subject to} \quad \ell \preceq x \preceq u$$

成为 constrained least squares (QP), 通常不再有同样简单的闭式解。



带先验的标定

若参考参数为 $\mathbf{x}_{\text{ref}}$ ，可使用

$$\min_{\mathbf{x}} \quad \frac{1}{2} \|\mathbf{A}\mathbf{x} - \mathbf{b}\|_2^2 + \frac{\lambda}{2} \|\mathbf{x} - \mathbf{x}_{\text{ref}}\|_2^2, \quad \lambda > 0.$$

Hessian 为

$$\mathbf{A}^\top \mathbf{A} + \lambda \mathbf{I} \succ 0,$$

故目标严格凸，并满足

$$(\mathbf{A}^\top \mathbf{A} + \lambda \mathbf{I}) \mathbf{x}^* = \mathbf{A}^\top \mathbf{b} + \lambda \mathbf{x}_{\text{ref}}.$$

数据拟合项与先验偏离项之间的权
衡；
 λ 控制正则化强度。

多面体之间的距离

$$\text{dist}(\mathcal{P}_1, \mathcal{P}_2) = \inf\{\|\mathbf{x}_1 - \mathbf{x}_2\|_2 \mid \mathbf{x}_1 \in \mathcal{P}_1, \mathbf{x}_2 \in \mathcal{P}_2\}.$$

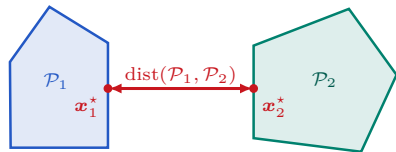
若

$$\mathcal{P}_1 = \{\mathbf{x} \mid \mathbf{A}_1\mathbf{x} \preceq \mathbf{b}_1\}, \quad \mathcal{P}_2 = \{\mathbf{x} \mid \mathbf{A}_2\mathbf{x} \preceq \mathbf{b}_2\},$$

则解 QP

$$\min_{\mathbf{x}_1, \mathbf{x}_2} \|\mathbf{x}_1 - \mathbf{x}_2\|_2^2 \quad \text{subject to} \quad \mathbf{A}_1\mathbf{x}_1 \preceq \mathbf{b}_1, \mathbf{A}_2\mathbf{x}_2 \preceq \mathbf{b}_2.$$

最优值为零当且仅当两多面体相交。

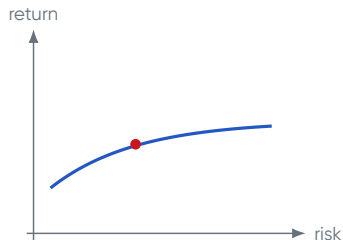


Markowitz 投资组合

令 x 为各资产的投资比例, $\bar{p}$ 为收益均值, $\Sigma \succeq 0$ 为收益协方差。

$$\begin{aligned} \min_x \quad & x^\top \Sigma x \\ \text{subject to} \quad & \bar{p}^\top x \geq r_{\min}, \\ & \mathbf{1}^\top x = 1, \quad x \succeq \mathbf{0}. \end{aligned}$$

$x^\top \Sigma x$ 是组合收益的方差; 该 QP 在保证最低平均收益与预算约束下最小化风险。



Boyd & Vandenberghe, Section 4.4.1

二阶锥规划

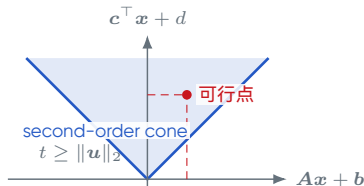
$$\begin{aligned} \min_x \quad & f^\top x \\ \text{subject to} \quad & \|A_i x + b_i\|_2 \leq c_i^\top x + d_i, \quad i = 1, \dots, m, \\ & Fx = g. \end{aligned}$$

约束

$$\|Ax + b\|_2 \leq c^\top x + d$$

等价于 $(Ax + b, c^\top x + d)$ 属于 second-order cone。

$A_i = 0$ 时退化为 LP；当 $c_i = 0$ 且 $d_i \geq 0$ 时，可平方为 QCQP；SOCP 比两者都更一般。



鲁棒线性约束

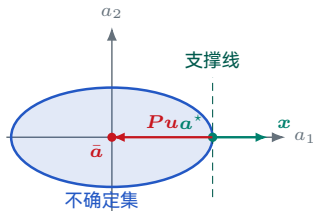
系数不确定但落在椭球中：

$$a = \bar{a} + Pu, \quad \|u\|_2 \leq 1.$$

要求 $a^\top x \leq b$ 对所有允许的 a 成立。当且仅当

$$\begin{aligned} \sup_{\|u\|_2 \leq 1} (\bar{a} + Pu)^\top x &= \bar{a}^\top x + \sup_{\|u\|_2 \leq 1} u^\top P^\top x \\ &= \bar{a}^\top x + \|P^\top x\|_2 \leq b. \end{aligned}$$

故 robust LP 可等价改写为 SOCP。



概率约束

若随机约束系数

$$\mathbf{a}_i \sim \mathcal{N}(\bar{\mathbf{a}}_i, \Sigma_i),$$

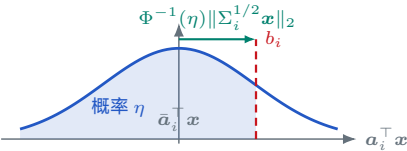
且要求

$$\text{prob}(\mathbf{a}_i^\top \mathbf{x} \leq b_i) \geq \eta, \quad \eta \geq 1/2,$$

则等价于

$$\bar{\mathbf{a}}_i^\top \mathbf{x} + \Phi^{-1}(\eta) \|\Sigma_i^{1/2} \mathbf{x}\|_2 \leq b_i.$$

右侧是二阶锥约束：给定置信水平的随机线性约束可在高斯模型下转为 SOCP。



单项式函数

在正交锥 $\mathbb{R}_{++}^n$ 上,

$$f(\boldsymbol{x}) = c x_1^{a_1} x_2^{a_2} \cdots x_n^{a_n}, \quad c > 0, \quad a_i \in \mathbb{R},$$

称为 monomial function。

这里的指数可以为负数或分数; 这与初等代数中“指数为非负整数”的 monomial 用法不同。

$$3x_1^{-1/2} x_2^2$$

在 $x_1 > 0, x_2 > 0$ 上
是 monomial。

正项式函数

monomial 的有限正和

$$f(\boldsymbol{x}) = \sum_{k=1}^K c_k x_1^{a_{1k}} \cdots x_n^{a_{nk}}, \quad c_k > 0,$$

称为 posynomial。

函数类	封闭运算
monomial	乘法、除法
posynomial	加法、乘法、正数乘
posynomial / monomial	仍为 posynomial

例如 $x_1^2 + 3x_1^{-1}x_2^{1/2}$ 是 posynomial。

几何规划的原始形式

$$\begin{aligned} \min_{\boldsymbol{x}} \quad & f_0(\boldsymbol{x}) \\ \text{subject to} \quad & f_i(\boldsymbol{x}) \leq 1, \quad i = 1, \dots, m, \\ & h_j(\boldsymbol{x}) = 1, \quad j = 1, \dots, p, \end{aligned}$$

其中 f_i 是 posynomial, h_j 是 monomial, 且隐含 $\boldsymbol{x} \succ \mathbf{0}$ 。

这称为 geometric program (GP) 的 posynomial form; 它通常不是原变量中的凸优化问题。

关键：经过对数变量替换与函数变换后，GP 成为凸问题。

GP 的等价规范化

GP 可以容纳更一般的写法:

$$f(\mathbf{x}) \leq h(\mathbf{x}) \iff \frac{f(\mathbf{x})}{h(\mathbf{x})} \leq 1,$$

其中 f 为 posynomial, h 为 monomial。

$$h_1(\mathbf{x}) = h_2(\mathbf{x}) \iff \frac{h_1(\mathbf{x})}{h_2(\mathbf{x})} = 1.$$

最大化非零 monomial 目标, 则最小化其 inverse monomial。
这些规则将工程表达式统一为 GP 的标准 posynomial form。



目标写为 inverse monomial 的最小化

对数变量替换

令

$$y_i = \log x_i, \quad x_i = e^{y_i}.$$

若 $f(x) = c \prod_i x_i^{a_i}$ 为 monomial, 则

$$f(e^y) = \exp(\mathbf{a}^\top \mathbf{y} + b), \quad b = \log c.$$

因此 monomial 在 log variables 中变为 affine function 的指数。单项结构由乘积变为加法。



GP 的凸形式

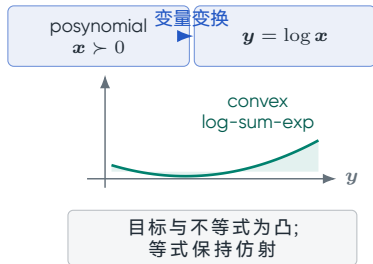
若 f 为 posynomial, 则

$$f(e^{\mathbf{y}}) = \sum_{k=1}^K \exp(\mathbf{a}_k^\top \mathbf{y} + b_k).$$

若单项式等式写成 $h_j(\mathbf{x}) = c_j \prod_{\ell} x_{\ell}^{a_{j\ell}} = 1$, 则对目标和约束取对数, 得到

$$\begin{aligned} \min_{\mathbf{y}} \quad & \log \sum_k e^{\mathbf{a}_{0k}^\top \mathbf{y} + b_{0k}} \\ \text{subject to} \quad & \log \sum_k e^{\mathbf{a}_{ik}^\top \mathbf{y} + b_{ik}} \leq 0, \\ & \mathbf{a}_j^\top \mathbf{y} + \log c_j = 0. \end{aligned}$$

log-sum-exp 是凸函数, 等式是仿射的; 这就是 GP 的 convex form。



悬臂梁设计

由 N 个单位长度矩形截面段组成的 cantilever beam, 编号从受力端向固定端依次为 $1, \dots, N$ 。设计变量为每段宽度 $w_i > 0$ 、高度 $h_i > 0$ 。

以总体积（重量）为目标，并限制截面尺寸、宽高比和最大应力：

$$\min_{w, h} \sum_{i=1}^N w_i h_i,$$

$$w_{\min} \leq w_i \leq w_{\max}, \quad h_{\min} \leq h_i \leq h_{\max},$$

$$S_{\min} \leq h_i/w_i \leq S_{\max}, \quad \frac{6iF}{w_i h_i^2} \leq \sigma_{\max}.$$



挠度约束

记 y_i 和 v_i 分别为第 i 段右端的挠度与斜率。由固定端条件 $v_{N+1} = y_{N+1} = 0$ ，从 $i = N$ 向 1 递推：

$$v_i = \frac{12(i-1/2)F}{Ew_i h_i^3} + v_{i+1},$$
$$y_i = \frac{6(i-1/3)F}{Ew_i h_i^3} + v_{i+1} + y_{i+1}.$$

每一步都是 monomial 与已有 posynomial 的和，所以 v_i, y_i 仍为 posynomial。特别地，受力端约束

$$y_1 \leq y_{\max}$$

是 posynomial inequality。

因此目标、尺寸约束、应力约束与挠度约束共同构成一个 GP。

固定端

$$v_{N+1} = y_{N+1} = 0$$



递推得到

$$v_N, y_N, \dots, v_1, y_1$$

从结构识别模型

类别	目标/约束结构	典型问题
LP	仿射	生产计划、网络流、Chebyshev 中心
QP/QCQP	凸二次 + 线性/二次约束	回归、距离、投资组合
SOC	仿射 + 二范数约束	鲁棒设计、概率约束
GP	posynomial/monomial, 经 log 变换	尺寸设计、缩放
一般凸问题	凸目标与凸不等式	正则化、估计

建模时首先辨认函数与约束的结构，再选择等价变换与适合的求解方法。



上海交通大学
SHANGHAI JIAO TONG UNIVERSITY

自然科学研究院
INSTITUTE OF NATURAL SCIENCES

Fin

